

**BRINGING HALO MODELING TOWARD  
PRECISION COSMOLOGY**

by

**Antonio S. Villarreal**

BS of Physics and Astronomy, University of Arizona, 2011

Submitted to the Graduate Faculty of  
the Kenneth P. Dietrich School of Arts and Sciences in partial  
fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH  
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Antonio S. Villarreal

It was defended on

July 20th, 2018

and approved by

Andrew Zentner, Department of Physics and Astronomy

Adam Leibovich, Department of Physics and Astronomy

Arthur Kosowsky, Department of Physics and Astronomy

Jeffrey Newman, Department of Physics and Astronomy

Hy Trac, Department of Physics, Carnegie Mellon University

Dissertation Director: Andrew Zentner, Department of Physics and Astronomy

# BRINGING HALO MODELING TOWARD PRECISION COSMOLOGY

Antonio S. Villarreal, PhD

University of Pittsburgh, 2018

The study of cosmology is entering an era of higher quality and higher quantity data that shifts us from limitations due to raw statistics of our data to limitations due to the accuracy of our underlying physical models. A specific set of models where this problem is prominent is the use of halo models to connect the invisible world of dark matter to the visible world of stars and galaxies. Most of these halo models make their predictions by taking the mass of dark matter halos as the sole parameter. The literature has demonstrated that secondary halo properties can have enhanced clustering compared to the general population of dark matter halos, in an effect that is referred to as “halo assembly bias.” Neglecting halo assembly bias from our models can result in severe biases. I add to the literature by making the first detailed study on the choice of halo definition on common measures of halo assembly bias. I utilize non-traditional halo definitions seeking a choice that minimizes the impact of environmental effects which may drive halo assembly bias. I find that halo assembly bias is a strong function of halo definition for the properties of halo concentration, halo shape, and halo spin. I demonstrate that the impact of halo redefinition is primarily caused by the changing host halo populations, as neighboring halos are demoted to substructure. I further show that these results are consistent with those of the “halo splashback radius”; however, halo splashback radius does not increase halo sizes sufficiently to remove halo assembly bias for most scales or masses. I discuss how these results give us insight to the relevant scales of what might be driving these relations and how they give a better understanding of galaxy formation and galaxy evolution. I conclude by laying out a course for the future with multiple paths to better understanding halo assembly bias and constraining how it impacts

the models as well as utilizing it as a probe of galaxy formation.

## TABLE OF CONTENTS

<b>1.0 INTRODUCTION</b>	1
<b>2.0 DARK MATTER AND HALO MODELING</b>	10
2.1 HALO MODELS	12
2.2 HALO ASSEMBLY BIAS	16
<b>3.0 METHODOLOGY</b>	21
3.1 SIMULATIONS	21
3.2 HALO FINDING	22
3.2.1 Halo Properties	23
3.2.1.1 Halo Masses and Radii	23
3.2.1.2 Halo Positions and Velocities	24
3.2.1.3 Halo Concentrations	24
3.2.1.4 Halo Shapes	29
3.2.1.5 Halo Spins	32
3.2.2 Halo Samples	33
3.2.2.1 Property Normalization	34
3.3 MEASURES OF ASSEMBLY BIAS	35
3.3.1 Correlation Function	35
3.3.2 Marked Correlation Function	36
<b>4.0 HALO ASSEMBLY BIAS IN DIEMER ET AL.</b>	38
4.1 CORRELATION FUNCTION RESULTS	38
4.2 MARKED CORRELATION FUNCTION RESULTS	39
4.2.1 Halo Concentration	39

4.2.2 Halo Shape . . . . .	42
4.2.3 Halo Spin . . . . .	44
<b>5.0 HALO REDEFINITION IN DIEMER ET AL.</b> . . . . .	47
5.1 MOTIVATION . . . . .	47
5.2 RESULTS . . . . .	48
5.2.1 Halo Concentration . . . . .	48
5.2.2 Halo Shape . . . . .	49
5.2.3 Halo Spin . . . . .	52
5.3 SUMMARY OF HALO CLUSTERING AND HALO REDEFINITION . . . . .	53
5.3.1 Halo Bias and Halo Redefinition . . . . .	53
5.3.2 How Does Halo Redefinition Help? . . . . .	57
5.3.3 How Does Halo Redefinition “Fail?” . . . . .	62
<b>6.0 HALO SPLASHBACK RADIUS AND ASSEMBLY BIAS</b> . . . . .	65
6.1 RESULTS . . . . .	65
6.1.1 Statistical Comparisons . . . . .	65
6.1.2 SPARTA Comparisons . . . . .	66
<b>7.0 TECHNICAL DISCUSSION</b> . . . . .	73
<b>8.0 CONCLUSIONS</b> . . . . .	78
<b>BIBLIOGRAPHY</b> . . . . .	83

## LIST OF TABLES

1	Minimum mass thresholds for our analyses. . . . .	33
---	---	----

## LIST OF FIGURES

1	The large scale distribution of dark matter in the Millennium Run. . . . .	4
2	The Tinker halo mass function. . . . .	14
3	The Tinker halo bias. . . . .	17
4	Halo assembly bias in the L0250 halo sample. . . . .	19
5	Halo properties as a function of halo definition in the L0250 catalog. . . . .	25
6	A toy model of halo redefinition. . . . .	26
7	The relationship between halo properties and halo masses. . . . .	30
8	A comparison of two halo concentration measures. . . . .	31
9	NFW concentration-dependent correlation functions. . . . .	40
10	NFW concentration marked correlation functions. . . . .	41
11	Velocity concentration marked correlation functions. . . . .	43
12	Shape marked correlation functions. . . . .	45
13	Spin marked correlation functions. . . . .	46
14	Concentration-dependant correlation functions as a function of halo definition. . . . .	50
15	NFW concentration marked correlation functions as a function of halo definition. . . . .	51
16	Velocity concentration marked correlation functions as a function of halo definition. . . . .	52
17	Shape marked correlation functions as a function of halo definition. . . . .	53
18	Spin marked correlation functions as a function of halo definition. . . . .	54
19	The relative halo bias for secondary halo properties as a function of mass. . . . .	56
20	Concentration marked correlation functions using fiducial halo properties and varying halo catalog with definition. . . . .	60



21	Velocity marked correlation functions using fiducial halo properties and varying halo catalog with definition. . . . .	61
22	A comparison of halo splashback radius to my best fit halo definition for assembly bias mitigation. . . . .	67
23	NFW concentration marked correlation functions for splashback radius definitions. . . . .	69
24	Shape marked correlation functions for splashback radius definitions. . . . .	70
25	Spin marked correlation functions for splashback radius definitions. . . . .	72

## 1.0 INTRODUCTION

Cosmology is the study of the evolution, content, and inevitable fate of the universe. It is a field of study that has gone through an enormity of changes over the last century; we have moved from being a field that was contemplating whether distant nebulae were truly extragalactic objects ([Hubble, 1929](#)) to one that has been rapidly narrowing in on a single model for explaining the universe with a mere six parameters (cite LCDM review papers + early LCDM papers here). In modern times, the rapid increase in data quality and quantity has allowed us to determine these parameters to greater and greater precision ([Riess et al., 1998](#); [Planck Collaboration et al., 2016](#)). Yet the model of the universe that we have developed has many unanswered questions that require further study. What is the dark energy necessary to explain the accelerating expansion of the universe at recent times ([Mortonson et al., 2014](#))? What is the dark matter that fits the extra gravitational force that we observe from the rotation curves of galaxies ([Garrett & Dūda, 2011](#))? Why do we have tensions between cosmological parameters such as the Hubble parameter at early times versus late times ([Freedman, 2017](#))? Do we see signals of modified gravity or does general relativity continue to hold in our observations ([Clifton et al., 2012](#))?

These unanswered questions have only served to raise the need for more advanced and precise experiments in a search for answers; and where our current data is potentially insufficient for teasing out the answers we desire, the future is very promising! There are many funded experiments that may shed light on some of these unanswered questions. The Large Synoptic Survey Telescope (LSST) will be able to take images of the entire night sky once every three nights and gather petabytes of data over ten years of operation ([LSST Science Collaboration et al., 2009](#); [LSST Dark Energy Science Collaboration, 2012](#)). The James Webb Space Telescope (JWST) ([Gardner et al., 2006](#)) and Wide Field Infrared Survey Tele-

scope (WFIRST) ([Spergel et al., 2015](#)) will gather high quality data without the troublesome properties of the Earth’s atmosphere, with image quality only limited by their optics. We are even starting to see the Universe in new ways that were previously just the limits of our imagination, with gravitational waves being observed multiple times by the ever-improving Laser Interferometry Gravitational-Wave Observatory (LIGO) ([Abbott et al., 2017](#)).

As data quality and quantity continues to increase within the field of observational cosmology, we find that the statistical limitations on our cosmological parameter constraints continue to decline. For some fields, this is a foray into a new era of precision cosmology; we now find ourselves in a situation where, rather than our ability to learn about our universe being limited by the raw statistics of our datasets, we are instead limited by how accurately we can model the underlying physics. This is a stark reversal of the past few decades, in which the observational errors far exceeded the theoretical! In fact, it has been shown that for a project such as LSST, the *systematic* error can lead to a significant bias in the cosmology parameters of interest ([Eifler et al., 2015](#)). This makes understanding of the underlying models that we are working with a critical problem of modern cosmology.

A valuable approach to handling these upcoming problems in modern cosmology is that of cosmological simulations. These computer simulations model the universe at early times and propagate these initial conditions forward to today. This allows us higher precision at smaller scales where most analytic approaches break down. In addition, it potentially allows us to battle a problem known as “cosmic variance.” In short, cosmic variance is the noise caused by random fluctuations in the early universe. Even an individual cosmological simulation can be limited by this fact, making direct interpretation difficult with regards to rare objects and events; this problem is most prominent for those simulations which are too computationally expensive to run many iterations. For most purposes though, simulations have the advantage of giving us a unique look at many different cosmological models and serves the form of experiment in lieu of being able to create an experiment in a laboratory. By drawing comparisons between these simulations and the observable universe, we can gain a better understanding of the world around us.

Roughly, these simulations break into two broad groups. The first is numerical

simulations only containing gravitational physics<sup>1</sup>. These simulations then trace the behavior of massive *particles* of collisionless dark matter from a set of initial conditions drawn from our understanding of the early universe to the future. These dark matter particles can be seen to cluster together under the force of gravity based on these early density fluctuations; commonly, we attempt to identify these clumps of dark matter as units known as “halos”.

The resulting clustering of dark matter can be seen to behave much like the visible matter in our observable universe. Dark matter halos merge hierarchically into large structures that match the “cosmic web”. It has even been shown that some of the largest structures, such as the Great Wall (Vogeley et al., 2004; Gott et al., 2005) feature in the Sloan Digital Sky Survey (SDSS) can be reproduced in sufficiently sized dark matter only simulations (Park et al., 2012). However, the ultimate limitation with dark matter only simulations is their inability to directly connect the structures of dark matter to the observable universe of stars and galaxies; for that step, some secondary modeling or assumptions need to be made.

This leads naturally to the other major approach: fully hydrodynamic simulations. These simulations use laborious calculations to solve the coupled differential equations that the baryonic particles of the universe obey. Computational advances have made this far more feasible in the last few years, allowing us to create galaxies that are comparable to those witnessed in the observed universe (Nelson et al., 2015; Schaye et al., 2015; Feng et al., 2016). There exist two primary limitations to this approach. The first is the fact that the complex calculations that must be solved still require state of the art computational resources and vast amounts of computing time. This has three major consequences: it makes it difficult to battle cosmic variance through large sample sizes, it makes it difficult to sample for many different sets of cosmological parameters, and it makes very large box sizes potentially unfeasible.

The second problem is perhaps more insidious: the limited mass and spatial resolution of these simulations forces us to consider “subgrid physics”. Consider for a moment that the masses of these particles are typically many orders of magnitude larger than that

---

<sup>1</sup>This is generally treated to be standard Newtonian gravity with an effective potential to account for expansion, although simulation frameworks utilizing modified gravity or including full general relativity are being explored in the field.

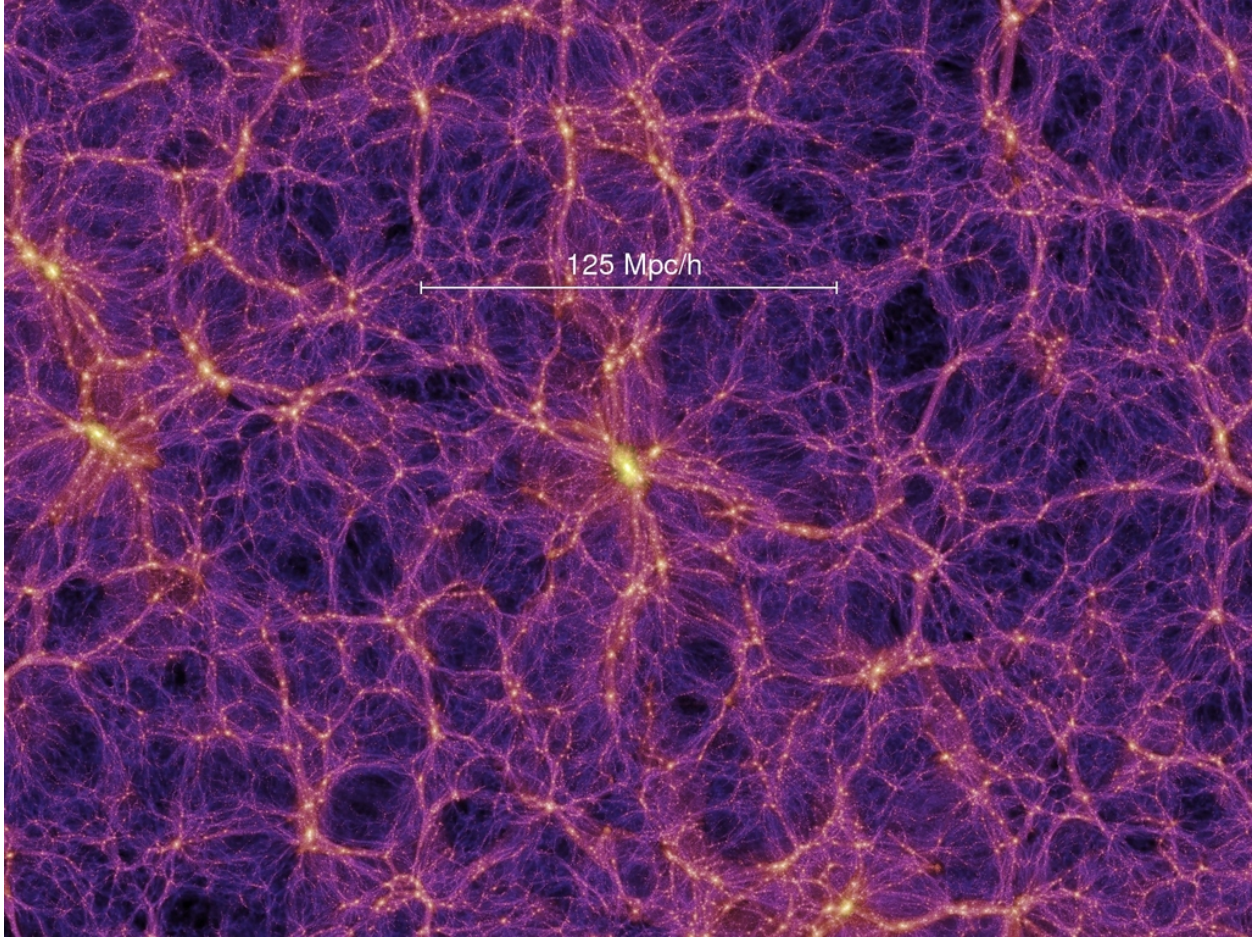


Figure 1 The large scale distribution of dark matter in the Millennium Run ([Springel et al., 2005](#)). Note the distinct “cosmic web” features, with many qualitative similarities to the Great Wall feature in SDSS.



of an individual star; however, the feedback driven by supernovae may have a significant impact on the formation of low mass galaxies (Katz, 1992; Smith et al., 2018, e.g.). In order to capture these effects which are smaller than the scales that we are able to simulate, we have to effectively add in the end result of the physics by hand. As of the time of this writing, there is no strong consensus in the field on how to include these subgrid models, nor is there a consensus on which subgrid physics can potentially be excluded or is most important, nor is there a consensus on how well different simulations should agree with each other considering different implementations. This leaves us with what is essentially a fine tuning problem; different choices of subgrid models and their input parameters may end up converging on the same resulting observables and may obscure the real galaxy formation physics or bias our cosmological predictions.

In lieu of being able to fully model all the necessary physics of galaxy formation, models must be developed in order to connect us from the density peaks of dark matter or our definitions of dark matter halos to the galaxies that we observe in the night sky. These range from more physically motivated to more empirically motivated, but typically are grouped together under the description of the “galaxy-halo connection.” The more physically motivated models include cases such as semi-analytic models (Benson, 2012; Croton et al., 2016, e.g.); these effectively use the knowledge that we have gained on the evolution of the density peaks over the history of the universe and galaxy formation. We add on a combination of analytic models and empirical relations for the underlying physics of star formation, gas cooling, feedback, and more. The more empirically motivated models focus more on the structures we refer to as dark matter halos and relating those directly to galaxies; for example, Halo Abundance Matching populates the most massive halos with the most massive galaxies (Hearin et al., 2013, 2014; Klypin et al., 2013; Desmond & Wechsler, 2017, e.g.), while the Halo Occupation Distribution takes a probabilistic view at populating individual dark matter halos with galaxies (Berlind & Weinberg, 2002, e.g.). These models, though fairly simplistic and seemingly naïve, reproduce many of the general clustering relations that we observe in the real universe, showcasing the strength of such techniques.

We group those models that create a direct link between dark matter halos to the observed galaxies as “halo models”. These models typically make two common assumptions

that are worth further consideration. The first is that all dark matter in the universe collapses into the halo structures that we observe. The second is that clustering and occupation of halo is a function of halo mass alone. But just how well do these two assumptions hold up in practice?

The former assumption deals more directly with a fundamental question I have left unanswered: “What is a dark matter halo?” The exact specifications as to how we choose to define a halo are left to be discussed in a further chapter, but there are broadly two classes of halo definition algorithms. It should be noted that implementations exist which are variants on these two general approaches<sup>2</sup> and many additional details exist on the specifics of individual halo finding algorithms.

The first is the “Friends-of-Friends” approach of halo finding. In this case a separation distance is chosen, called a linking length, by which dark matter particles can be separated by and still be considered a single halo object. Dark matter particles are then linked together in chains by this distance or smaller in order to generate large halos of irregular size and shape. Smaller halos, as might be expected with hierarchical formation, can be found by looking at smaller linking lengths within an existing group of dark matter particles. The resulting dark matter halos will have a halo mass determined by the number of linked particles (and the mass of the particle) and may have a considerably non-spherical shape. Note that halos generated using this method are highly sensitive to the choice of linking length; one can imagine a linking length which connects all particles as a single halo or entirely misses structure on the fringes of the halo. A known problem with this method is the identification of substructures within larger halos. This algorithm can easily artificially identify overdensities which are not gravitationally bound as the linking length is reduced. As these objects are merely transient structures, this may bias halo statistics, particularly for the least massive halos.

The second is the “Spherical Overdensity” approach of halo finding. Peaks are identified in the underlying matter density field; for each peak above a threshold height, a spherical halo is created such that the encompassed mass results in an average density that is a multiple,  $\Delta$ , of the mean matter density or critical density of the simulated universe. In

---

<sup>2</sup>Including the ROCKSTAR method discussed in our methodology.

this case, substructure is identified by having a halo center within the radius of a larger halo. The main advantage to this technique is a direct link to models of spherical collapse, where analytic calculations can be computed. For an Einstein-de Sitter universe, the resulting value of  $\Delta$  for a gravitationally bound object is 178 with respect to the critical density of the universe (Mo et al., 2010, e.g.). A common halo definition used in the literature chooses the different value of  $\Delta = 200$  with respect to the mean matter density of the simulation (Miyatake et al., 2016; Paranjape et al., 2018, e.g.). This methodology carries the advantage of being easily linked to the multitude of properties associated with spherical halos, including that of the Navarro-Frenk-White (NFW) (Navarro et al., 1997) profile for the distribution of halo mass.

Note now that neither halo definition adequately addresses the question of “does all dark matter mass reside in dark matter halos”. Work has been done to demonstrate that potentially considerable amounts of matter reside outside of spherically defined halos (van Daalen & Schaye, 2015). One interpretation is that significant matter density exists inside of filamentary structures connecting halos; another is that traditional spherical halo definitions simply miss a large amount of matter on scales larger than the virial radius. But while a spherical overdensity method may miss matter, a FOF approach runs the risk of linking together halos in an artificial manner.

I note the arbitrary nature of the cutoff of what defines a halo as a particular point of interest to us. While the Spherical Overdensity method is more physically motivated, note that the fiducial choice of  $\Delta$  is considerably different from the analytic result of an Einstein-de Sitter universe (a value of  $\Delta = 178$ ) or even the  $\Lambda$ CDM universe that we reside in (a value of  $\Delta = 337$ ). Changing this value of  $\Delta$  can considerably change the definition of the halo radius and thus the contained dark matter and the measured distribution; this could impact any number of halo properties of interest. In fact, different analysis groups in the field choose to look at different definitions simply due to matters of convenience and little attention has gone into the careful analysis of how halo definition matters. This makes it exceedingly difficult to compare the results of studies at the level of precision necessary to address future high-quality data. I will refer back to this problem to motivate our methodology of halo redefinition in later chapters.



I now draw attention to the second assumption: that at fixed halo mass, halos are self-similar (with some scatter). Setting aside the matter that the measured halo mass is a function of your choice of halo definition, it has been shown in the literature extensively that this assumption is false. In fact, the clustering of halos has been shown to depend on halo formation time (Gao et al., 2005; Harker et al., 2006; Wechsler et al., 2006; Gao & White, 2007; Croton et al., 2007; Zentner, 2007; Dalal et al., 2008; Li et al., 2008; Lacerna & Padilla, 2011), halo concentration (Wechsler et al., 2006; Faltenbacher & White, 2010), and other halo properties (Bett et al., 2007; Hahn et al., 2007b,a, 2009; Faltenbacher & White, 2010; Hester & Tasitsiomi, 2010; Lacerna & Padilla, 2012; van Daalen et al., 2012; Fisher & Faltenbacher, 2016; Sunayama et al., 2016; Chaves-Montero et al., 2016). But with such a wealth of the literature pointing toward this seeming breakdown in the halo model, why even persist with such a method?

One reason is that it is clear that halo mass is the dominant variable in determining halo clustering and occupation (Efstathiou et al., 1990; Bond et al., 1991; Cole & Lacey, 1996; Zentner, 2007). Another reason is that the halo model is important in providing us a physical intuition for how to connect the underlying matter power spectrum of the early universe to the observed universe around us. The hierarchical formation of structure in dark matter simulations has motivated much of our current understanding of how galaxies form. Understanding the connection between dark matter and galaxies helps us to gain further insight, even using empirical models with the currently identified flaw.

However, ignoring this breakdown of our assumptions could lead to unexpected consequences. As such, it is important to study the impact of this; the most common method is to look at the “halo assembly bias”, measured as the difference in clustering that is demonstrated as a function of a secondary halo property. I study the impact of halo assembly bias as a function of halo definition (and necessarily halo size in terms of the halo radius) for halo concentration, halo spin, and halo shape. I extend the choice of halo definition outside the bounds of the customary definitions in order to explore just how much impact this can have on measure of halo assembly bias and as an attempt to find a potentially more physically motivated halo definition that minimizes assembly bias. This provides the first detailed look in the literature on how our choice of halo definition can drastically change how we identify

halo assembly bias and how it may be impacting our predictions. An additional benefit is a firm understanding of the relationship between measured halo bias and halo definition allows us to better compare the results of previous authors from each other.

A halo definition free from assembly bias would be applicable to all of our previous halo models without needing to involve additional parameterizations. I motivate drastically different choices of the overdensity parameter  $\Delta$  based on the fairly arbitrary choices that have become the fiducial definitions in the literature and explore these for the first time. I focus on halo definitions that yield significantly larger halo radii than traditional choices in the literature, in the hope of encompassing changes driven by the local environment. I also examine the physically motivated “splashback” definition, in which infalling material clumps in caustics. I look at this both in a broad statistical sense using a fitting function, as well as with the more computationally expensive methods of assigning splashback radius to each halo individually.

In the remainder of this thesis, I go into further detail on each of these topics. In Chapter 2, I look into further detail on the definition of common halo models and their application in the field. I further demonstrate how halo assembly bias is defined and has been measured in the existing literature. In Chapter 3, I discuss the simulations that I utilize in my analysis and the **ROCKSTAR** halo finder used in my halo identification. I additionally discuss two common measures of halo assembly bias: the correlation function (or two-point autocorrelation function) and the marked correlation function. In Chapter 4, I show our detection of halo assembly bias in my simulation suite as a function of four distinct halo properties. In Chapter 5, I examine how changing halo definition can directly impact the measure of halo assembly bias and how this, itself, is dependent on which halo property is examined. In Chapter 6, I study if the “splashback radius” explains the differences in halo assembly bias as a function of halo redefinition. This is also tested as a possible more physically motivated definition that may be free of these problems. In Chapter 7, I narrow down on the technical results and how they compare to existing results in the literature. In Chapter 8, I discuss the larger context of those results within cosmology and what my future aims are for developing a better understanding of these models for cosmological predictions.

## 2.0 DARK MATTER AND HALO MODELING

In the previous chapter, I briefly mentioned one of the fundamental mysteries of modern cosmology: “What is dark matter?” As this work focuses on the nature of how dark matter structures evolve in the universe and how we can utilize our knowledge of dark matter clustering and occupation in order to push cosmology forward, this enigma merits close examination. The history of this question can ultimately be traced back to the early work by Fritz Zwicky, who identified that the velocity dispersions of seven galaxies in the Coma cluster implied a gravitational mass that was significantly more than the observed stellar mass (Zwicky, 1933). This extra matter received the name “dunkle Materie”, rather than being simply described as something along the lines of missing matter. This work soon expanded to the Virgo cluster (Smith, 1936), though with significant amounts of room for error due to the methodologies of the time.

The following years would unveil more hints of the presence of dark matter. Galaxy rotation curves were measured to higher precision and larger distances by independent groups within the literature (Freeman, 1970; Bosma, 1978; Rubin et al., 1980). The general result was the measurement of a flattened rotation curve out to large distances from the center of the galaxy; this observation is exceedingly difficult to reconcile with the observed mass, leading to two competing conclusions. The first conclusion is that the inability to resolve the difference between the observed mass and gravitational interaction implies a breakdown in our understanding of gravity; this motivates the field of modified gravity within the literature as one starting point. The other conclusion is that there is some source of mass invisible to the electromagnetic spectrum: “dark matter”. Turner et al. (1984) additionally demonstrate that structure cannot develop fast enough to produce the galaxies that we observe today with baryonic matter alone; one solution suggested is the existence of a relic particle such as dark

matter.

I choose to examine dark matter in a phenomenological sense throughout this work. Specifically, I am less interested in dark matter on the level of individual particles<sup>1</sup>, but rather I focus on the bulk behavior of the dark matter. With regards to this, I consider the most common model of *cold dark matter*. These particles are **cold** in that their velocities are significantly smaller than the speed of light and **dark** in the sense that they are not observed electromagnetically. I note this is not the only possible model: warm dark matter has enjoyed some popularity due to the ability to help explain the density profiles of dark matter that are inferred (Bode et al., 2001; Götz & Sommer-Larsen, 2003) and some annihilating or interacting dark matter models can produce Standard Model particles that could potentially be observed (Bertone et al., 2005, for a thorough review of the particle approach). As of the time of writing, constraints tend to favor the CDM model (Inoue et al., 2015; Yèche et al., 2017); it does remain an active discussion within the literature, however.

For the paradigm of cold dark matter, there has been extensive work carried out in the literature to study the behavior with computer simulations. This work was carried out with analog computers for galaxies as early as 1941, where the flux of light bulbs was used to measure the gravitational potential (Holmberg, 1941). From there, simulations were made digital alongside the development of the cold dark matter model itself. The rapid acceleration of computational power, known as Moore’s Law, has lead us to an era where we carry out simulations of unprecedented resolution year after year. For models with only dark matter and their gravitational interactions in consideration, we have the Millennium Simulation (Springel et al., 2005) and the Bolshoi Cosmological Simulation (Klypin et al., 2011). Far more complicated simulations take into account baryonic physics such as star formation and feedback from Active Galactic Nuclei (AGN) in simulations such as the Illustris Simulation (Vogelsberger et al., 2014) or the BLUETIDES Simulation (Feng et al., 2016). However, this field of hydrodynamic simulations still requires significant progress to be made to converge to agreements between simulations (Scannapieco et al., 2012). As such, I focus on the results of dark matter only simulations and their interpretation.

---

<sup>1</sup>If it is indeed a particle...

## 2.1 HALO MODELS

Consider the endpoint of a numerical, dark matter only simulation: a collection of particle positions and velocities. This is a far cry from the observations that we make of the galaxies on a night sky! The ways we can make the connection between the two range from the semianalytic to the phenomenological, but most take the following broad form:

1. Identify collections of dark matter particles referred to as “halos”.
2. Calculate properties of the dark matter halos.
3. Use these properties to populate dark matter halos with galaxies.

The first step requires us to identify some collection of dark matter as “halos” A major motivation for this approach is the ease at connecting this to early analytic approaches. One can make the assumption that non-linear objects form from a spherically symmetric collapse starting from density perturbations in the early universe ([Gunn & Gott, 1972](#)). For an Einstein-de Sitter universe, this calculation can be carried out analytically, resulting in determining that a virialized object in the universe has a density equal to

$$\Delta_{\text{vir}} \equiv \frac{9\pi^2}{16} \times 8 \times 4 = 18\pi^2 \approx 178 \quad (2.1)$$

times the background density of the universe at the time of virialization ([Cooray & Sheth, 2002](#), for a derivation). This calculation can be carried out for the  $\Lambda$ CDM universe and yields  $\Delta_{\text{vir}} \approx 337$ . The idea of following spherical collapse was developed into the formalism of [Press & Schechter \(1974\)](#) and extended ([Bond et al., 1991](#); [Bower, 1991](#); [Kauffmann et al., 1993](#); [Lacey & Cole, 1993](#)) to allow the prediction for the clustering of dark matter halos and the evolution of the halo mass function, a measure of the number density of halos of a given mass in the universe.

These results motivate one of the common methods of identifying halos in simulations: “spherical overdensity” (SO) identification. One of the most prominent early uses of this method rests on [Lacey & Cole \(1994\)](#), where the method was used as a numerical test as to the accuracy of their Extended Press-Schechter formalism from [Lacey & Cole \(1993\)](#). These halo finders comb through the density field, identify local density peaks that exceed

this required limit to be virialized, and creates a spherical halo around these points. In particular, the [Klypin & Holtzman \(1997\)](#) algorithm to carry out this calculation remains in extensive use to this day.

A complementary approach is the “Friends-of-Friends” (FOF) method of halo finding, which chooses to identify halos through a linking length parameter,  $b$ . This tunable parameter serves as a length by which particles in a simulation are connected together; all connected particles constitute a single halo ([Davis et al., 1985](#)). While this approach has a weaker connection to the underlying analytical theory, it serves as a powerful approach for dealing with the real data which is both highly non-spherical and asymmetric.

The overwhelming majority of halo finding algorithms utilized by the field build off the work of these two approaches. For a specific example, [Behroozi et al. \(2013b\)](#) provides the ROCKSTAR halo finding algorithm. This approach identifies the equivalent to FOF groups inside of a 6D phase space before calculating halo properties in spherical overdensity apertures. Overall, while many halo finders struggle to identify substructure, these phase space oriented halo finders are able to resolve substructure down to a small number of particles ([Knebe et al., 2011](#)), though a larger number of particles is often needed to accurately determine the properties of these smaller halos ([Onions et al., 2012](#)).

A particular quantity of interest to anybody carrying out calculations from dark matter halos is that of the halo mass function, or the number of halos of a given mass in the universe. While analytic methods do exist of estimating this function ([Linke et al., 2017](#)), these functions are typically determined from simulation. By far the most common halo mass function in the literature is that of [Tinker et al. \(2008\)](#), reproduced as Fig. 2. Note the shape of the halo mass function; there exist many more halos of small mass than halos of large mass. As the universe advances through time, these small halos merge to form more massive halos, pushing this distribution toward higher mass. Note also that very high mass objects become increasingly rare, especially past the visible turnover in the halo mass function. Even more modern advances in the study of the halo mass function remain rooted in the study of simulation; it is now popular to generate emulators which are able to interpolate between the results of many halo simulation boxes in order to generate even more robust estimates of the halo mass function as a function of cosmology ([McClintock et al., 2018](#)).

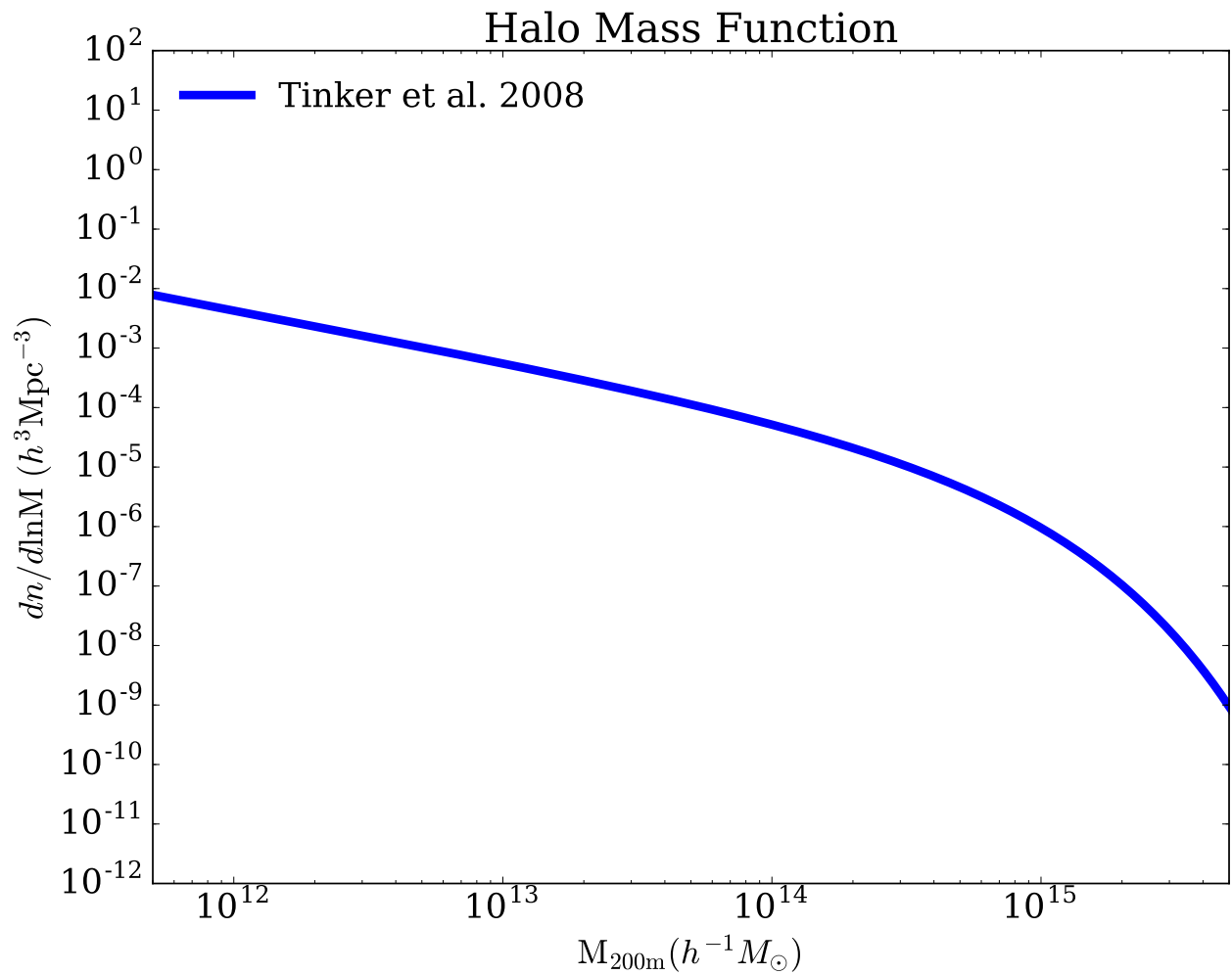


Figure 2 The halo mass function from [Tinker et al. \(2008\)](#) as the number density of galaxies in bins of logarithmic mass. Note that the function monotonically declines, with the beginning of a turn-over to a more rapid decline visible at the highest mass bins.

Once a dark matter halo has been defined, there are a number of properties that are of interest for studying. The first (and potentially one of the more varying) property to be considered is that of the halo mass. Once a decision has been made on which particles belong in a halo, it is a simple matter of addition to determine the total mass of dark matter particles contained in the halo. It is often beneficial to assign a “halo radius” to a given object; for a spherical overdensity halo a common choice is to define the radius as the scale such that the average density within a sphere of that radius exceeds some multiple of the background. As the halo size increases, the expectation is for the average density to decrease, leading to a threshold size. Other properties include: “halo concentration”, which provides a measure of how a characteristic scale in the density profile of dark matter compares to the total size of the halo; “halo shape”, which is a measure of how much the halo deviates from the assumption of a perfect sphere; and “halo spin”, which measures how much energy is stored in angular momentum for the halo, often with very different normalization between chosen definitions. These properties are all able to be directly pulled from the simulation data at the particle level and serve as our input parameters to various models.

With our dark matter halo properties, we can now move to populate these dark matter halos with galaxies and get a sense of how both halos and the objects that they live in cluster together. This connection is motivated by the work of [White & Rees \(1978\)](#), which proposed that dark matter halos are occupied by galaxies as luminous cores. The simplest property that one might connect from halo to galaxies is that of halo mass. The intuition behind this is fairly straightforward; a sufficiently massive halo will have a gravitational potential that will attract more matter. This will cause the halo to effectively evolve faster (undergoing more mergers at earlier times) and gather more material in the early universe (increasing the baryonic mass within). In fact, one can observe this in simulation fairly straightforwardly; the clustering of halos at a given mass can be enhanced from a uniform distribution in what is referred to as the halo bias. I show this in [Fig. 3](#) using the fitting function provided in [Tinker et al. \(2010\)](#); here the halo bias is determined from simulation as

$$b^2(k) = P_h(k)/P_{\text{lin}}(k), \quad (2.2)$$

where the bias is a ratio of the halo power spectrum to the linear dark matter power spectrum.



As the halo bias is a very strong function of the halo mass and we expect that galaxies trace the underlying dark matter population, we can infer the following: the most clustered galaxies tend to populate the most massive dark matter halos. This idea allows us to link together our understanding of dark matter halos and their clustering to the study of galaxy formation and galaxy clustering in a fundamental way.

This motivates further methods such as the Halo Occupation Distribution (HOD) (Mo & White, 1996; Berlind et al., 2003; Kravtsov et al., 2004; Zheng et al., 2005, 2007) to that of Halo Abundance Matching (Hearin et al., 2013, 2014; Klypin et al., 2013; Desmond & Wechsler, 2017). These models can often match observables such as the galaxy-galaxy clustering quite well and serve as a valuable first step for gaining insight into both the formation of galaxies and their connection to the underlying dark matter halos. As a result of this strong connection, many models have been generated that assume that *only* halo mass matters for the halo-galaxy connection or even halo-halo clustering. This assumption is bolstered by the fact that many of the halo properties discussed are strongly correlated with halo mass. The breakdown of this assumption is often referred to as “halo assembly bias” (when referring to halo-halo clustering) or “galaxy assembly bias” (when referring to how this might impact the halo-galaxy connection). As of the time of writing, the topic of how failure to address assembly bias impacts various observation remains a source of heated debate. In the next chapter, I will detail what progress has been made and where the field currently stands.

## 2.2 HALO ASSEMBLY BIAS

In the previous section, I discussed the matter of halo models that are focused on halo mass as the most important and (often) only parameter. Early work by Gao et al. (2005) demonstrated that the clustering of dark matter halos exhibited dependence on an additional property: halo age. This sparked the study of the excess clustering of dark matter halos as a result of a secondary halo property, typically referred to as “halo assembly bias.” Halo assembly bias has been detected in association with halo formation time (Harker et al.,

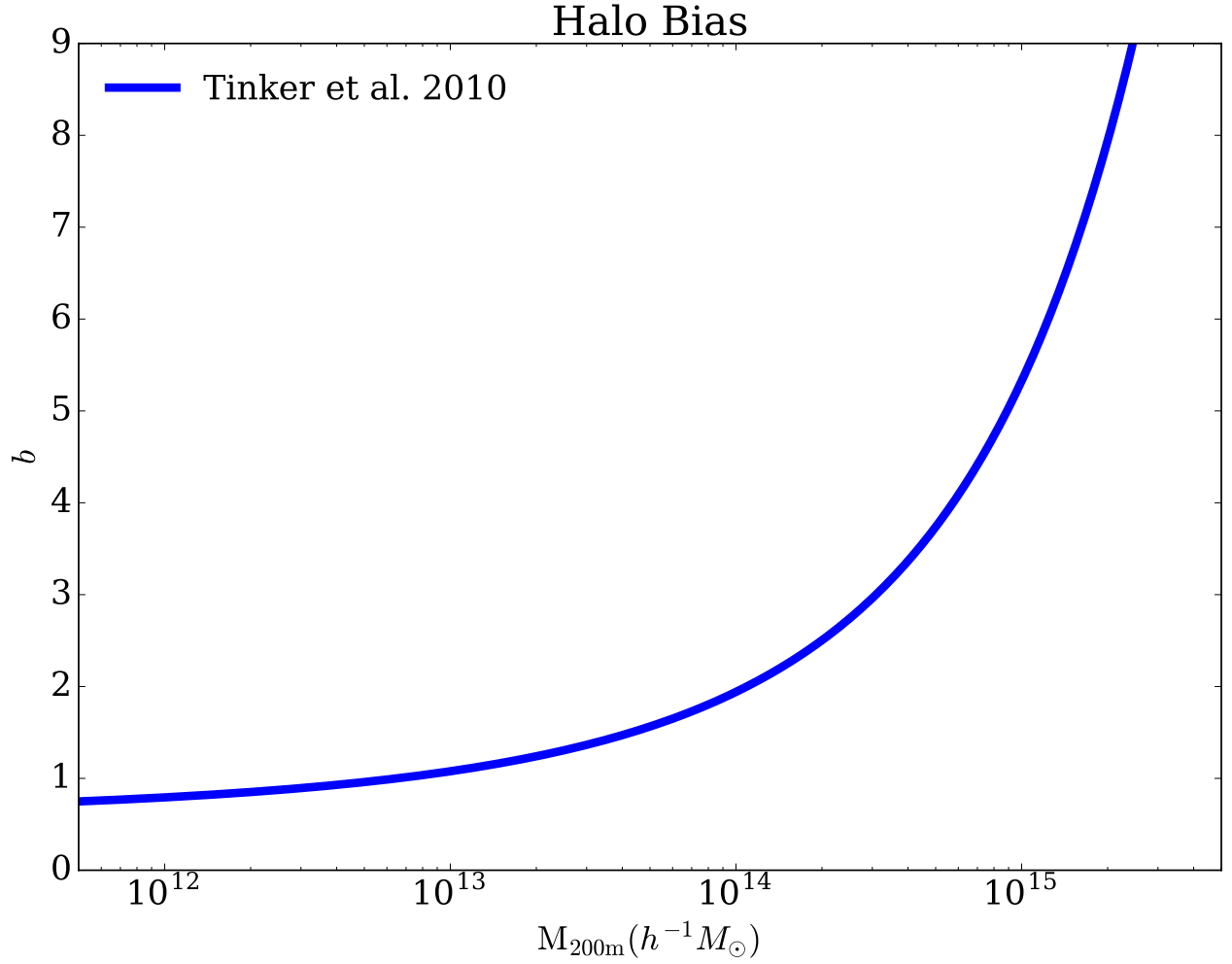


Figure 3 The large scale halo bias as measured by the ratio  $(P_h/P_{lin})^{1/2}$  using the [Tinker et al. \(2010\)](#) halo bias fitting function, assuming a  $\Delta = 200m$  halo definition.

2006; Wechsler et al., 2006; Gao & White, 2007; Croton et al., 2007; Zentner, 2007; Dalal et al., 2008; Li et al., 2008; Lacerna & Padilla, 2011), halo concentration (Wechsler et al., 2006; Faltenbacher & White, 2010), and other halo properties (Bett et al., 2007; Hahn et al., 2007b,a, 2009; Faltenbacher & White, 2010; Hester & Tasitsiomi, 2010; Lacerna & Padilla, 2012; van Daalen et al., 2012; Fisher & Faltenbacher, 2016; Sunayama et al., 2016; Chaves-Montero et al., 2016). For an example of what halo concentration driven assembly bias can look like, I present Fig. 4, which demonstrates halo assembly bias when the halo sample is split in half based on high or low concentration at fixed halo mass. Those halos with higher concentrations at fixed mass have enhanced clustering, while those halos of lower concentrations at fixed mass have decreased clustering.

The impact halo assembly bias has on existing results in the literature remains unknown. Gallart et al. (2015) suggests that the difference between dwarf galaxy types can be linked to the impact of halo assembly bias. Mao et al. (2015b) notes that halo assembly bias can have an impact on subhalo abundances that should be observable in future deep surveys, allowing for the possibility of halo assembly bias being verified observationally. Wang et al. (2018) demonstrates that while halo mass is the primary driving force between galaxy quenching in their sample, there remains a residual environmental dependence which can be modeled out as halo assembly bias. The ability of halo assembly bias, which is unsurprisingly difficult to determine through direct observation, to impact galaxy formation and observation is more than ample motivation to study the impact of it in depth. Yet even more motivation exists! It has been demonstrated that as we move to higher quality, higher quantity, and deeper cosmological surveys, we begin to enter a regime in which not accounting for halo assembly bias will hamper your ability to measure cosmological parameters correctly: in particular dark energy (e.g., Croton et al., 2007; Wu et al., 2008). A better understanding of halo assembly bias will be critical to being able to mitigate any potential systematic biases that this might introduce.

In the next chapter, I will detail our own procedure of detecting halo assembly bias. Our prescription is similar to that of Wechsler et al. (2006), with modifications to removing the underlying secondary halo property to halo mass relations. I go into detail on the dark matter only simulations that were utilized in this analysis. I discuss our choice of halo finding

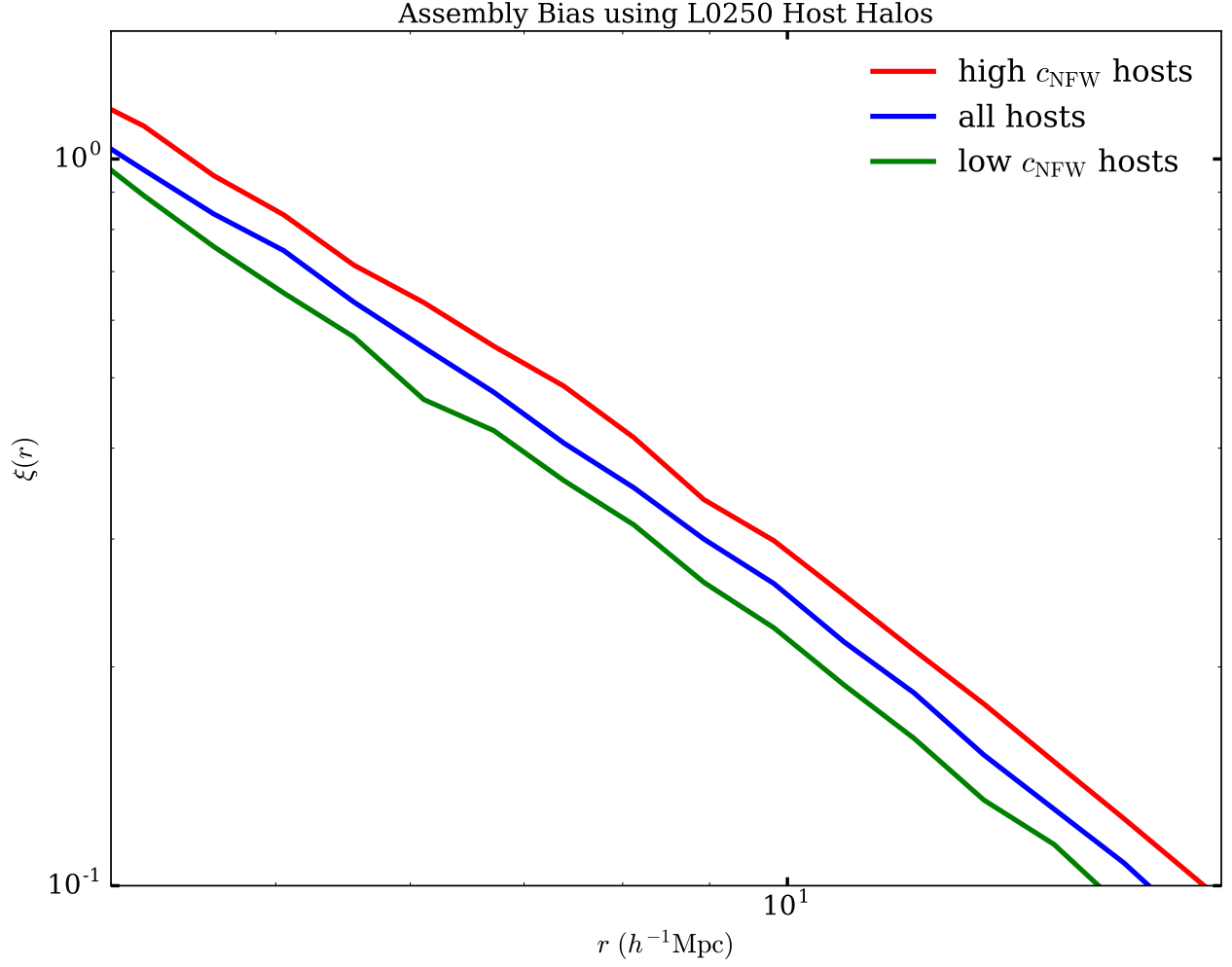


Figure 4 The 2-point correlation function measured for host halos in the L0250 simulation. The blue line contains all host halos in the sample, while the red (green) line contains the 50% highest (lowest) concentration halos at fixed halo mass.

algorithm and cuts on our overall halo sample to assure that properties are well resolved. I finally discuss the actual mathematics of halo assembly bias, in particular the less familiar “Marked Correlation Function.”

### 3.0 METHODOLOGY

In this chapter I shall discuss in detail the methods utilized in our analysis. In Section 3.1, I cover the [Diemer & Kravtsov \(2015\)](#) simulations that have been provided to me for analysis and their cosmology in detail. In Section 3.2, I describe the ROCKSTAR halo finding algorithm that I run on our simulations for different choices of halo definition. I further describe the various halo properties that are utilized in the analysis and how these properties are determined by ROCKSTAR. In Section 3.3, I discuss how I measure halo assembly bias, both with more traditional correlation functions and the potentially more powerful marked correlation functions.

#### 3.1 SIMULATIONS

I make use of three cosmological N-body simulations in order to study the impacts of halo assembly bias throughout this work. These simulations are a subsample of the [Diemer & Kravtsov \(2015\)](#) simulations that have been performed using cosmological parameters of  $\Omega_M = 0.32$ ,  $\Omega_\Lambda = 0.68$ ,  $h_0 = 0.67$ ,  $\sigma_8 = 0.834$ , and  $n_s = 0.9624$ . This cosmology was chosen due to its close resemblance to the Planck best-fit cosmology as presented in [Planck Collaboration et al. \(2014\)](#). The initial matter power spectrum was generated using the CAMB code for this set of cosmological parameters ([Lewis & Bridle, 2002](#)). Each simulation further has initial conditions generated using the second-order Lagrangian perturbation theory code 2LTPC<sub>IC</sub> ([Crocce et al., 2006](#)) from a starting redshift of  $z = 49$ . This initial redshift is sufficiently high to avoid potential transient effects that are present with simpler approximations at low initial redshifts ([Crocce et al., 2006](#)). In addition, I focus on the  $z = 0$  redshift snap-

shots for this study. The force solving algorithm used in these simulations is the **GADGET2** code from [Springel et al. \(2005\)](#).

Together, the simulations cover box sizes of 125, 250, and 500  $h^{-1}\text{Mpc}$  respectively; I refer to them throughout this work as L0125, L0250, and L0500 from now on. Each contains  $1024^3$  particles, leading to particle masses of  $1.6 \times 10^8$ ,  $1.3 \times 10^9$ , and  $1.0 \times 10^{10} h^{-1} M_{\odot}$  in each simulation volume. For purposes of this study, the key difference between simulation boxes is the mass resolution available. L0125 contains few massive halos, but a large number of small mass halos; L0500 is incapable of resolving the least massive halos, but has many more of the most massive halos. This allows me to do a comprehensive study of the potential impacts of halo assembly bias as a function of halo mass. Each simulation also contains a slightly different force softening length as a result of different particle loading; 2.4, 5.4, and  $14 h^{-1}\text{kpc}$  from smallest box size to largest. I note that these lengths are significantly smaller than the physical scales of interest.

### 3.2 HALO FINDING

The matter of identifying dark matter halos is a very non-trivial part of this analysis. It is clear that while halo finders asymptote to agreement on most statistical properties, the specifics between a FOF halo finder versus a Spherical Overdensity halo finder can be considerably different. I identify dark matter halos in the simulation snapshots using the **ROCKSTAR** halo finder from [Behroozi et al. \(2013a\)](#), which uses a variant approach to the latter. Schematically, the approach of **ROCKSTAR** is as follows:

1. Divide the simulation volume into rough Friends-of-Friends groups, using some set linking length,  $b$ .
2. Inside of each FOF group, normalize member positions and velocities by the group position and velocity.
3. Adaptively link particles together in the 6D phase-space such that 70% of particles are linked together into smaller groups.
4. Repeat this to find more levels of substructure until new levels cannot be found.

5. Place seed halos inside the lowest level of substructure and assign nearby particles in phase-space membership to these halos.
6. Remove any unbound particles and calculate halo properties using their member particles.

This particular algorithm has several advantages over the more basic algorithms discussed previously in the text. The first is that phase-space algorithms are exceptionally good at being able to identify substructures (Knebe et al., 2011). The second is that this analysis combines the easily interpreted SO halo properties with the more robust halo identification of FOF halo finding. Note that I have chosen the linking length for our analysis as a relatively large linking length of 0.4. I have varied this choice and found that it makes no qualitative difference in the following analysis.

### 3.2.1 Halo Properties

As briefly mentioned above, ROCKSTAR calculates halo properties primarily through spherical overdensity methods. This addresses an issue in which FOF groups will often be grouped together well in advance of SO halos (Klypin et al., 2011). I detail below the specific procedure by which ROCKSTAR handles each given property that is of interest in this work.

**3.2.1.1 Halo Masses and Radii** The definition of halo mass, on the exterior, is a fairly straightforward calculation. One needs to merely add up the total number of particles belonging to a dark matter halo; since one knows the mass of each individual particle, the mass of the dark matter halo is trivially known. However, the *size* of a dark matter halo is not necessarily a simple choice to make.

In practice, for SO halos, we determine the halo size with respect to a multiple of the mean background density,  $\rho_m$ :

$$\bar{\rho}(R_\Delta) = \Delta\rho_m. \quad (3.1)$$

Here,  $\Delta$  is a (typically) integer multiple and  $\bar{\rho}(R)$  is the mean density inside a sphere of given radius,  $R$ . I choose to mark our numerical choices of  $\Delta$  with an “m” if the value is with



respect to the mean background density and a “c” if it is with respect to the critical density. This choice of  $\Delta$  directly relates to earlier discussion of spherical collapse and has lead to a common choice in the literature of  $\Delta = 200m$  for most studies related to halo assembly bias. However, this decision is not universal; ranges of values from  $\Delta \approx 178m$  to  $\Delta \approx 340m$  are fairly common in the literature. Even the choice of choosing the mean background density is not without consequence, as some analysis is done using the critical density instead (roughly resulting in  $\Delta = 200c \approx 625m$ ). Another common choice is the virial overdensity (Bryan & Norman, 1998), which comes out to  $\Delta = 100c \approx 300m$ . Outside of the field of assembly bias, even larger values of  $\Delta$  are chosen to focus on the innermost regions of galaxy clusters, sometimes reaching values as high as  $\Delta = 625m$ . Once a halo size is defined using this overdensity definition, calculating the halo mass is straightforward; Fig. 5 demonstrates how the choice of halo size impacts the determination of halo mass.

I provide a toy diagram in Fig. 6 to provide an intuition on how the choice of halo definition could have significant impact on halo mass. In this cartoon, the solid lines correspond to a traditional halo definition with  $\Delta = 200m$  and the dashed lines correspond to a significantly extended halo of  $\Delta = 20m$ . Note that the drastically increased halo definition will lead to a halos C and D to cease to be distinct “host” halos in the larger halo definition; their inclusion in the halo mass calculation could considerably change this and other properties.

**3.2.1.2 Halo Positions and Velocities** ROCKSTAR follows an approach by which the center of the halo is measured by following the density peaks rather than through averaging. It first determines the inner substructure of the halo such that the expected Poisson error,  $\sigma_x/\sqrt{N}$  is minimized. This central substructure then has its particles utilized to calculate the halo center. The velocity of the halo is then calculated averaging the velocities of those particles within the innermost 10% of the halo radius.

**3.2.1.3 Halo Concentrations** I investigate two distinct definitions of halo concentration. Halo concentration comes out of the NFW profile that describes the spatial distribution of dark matter halos in simulation. While this is known not to match up with observations

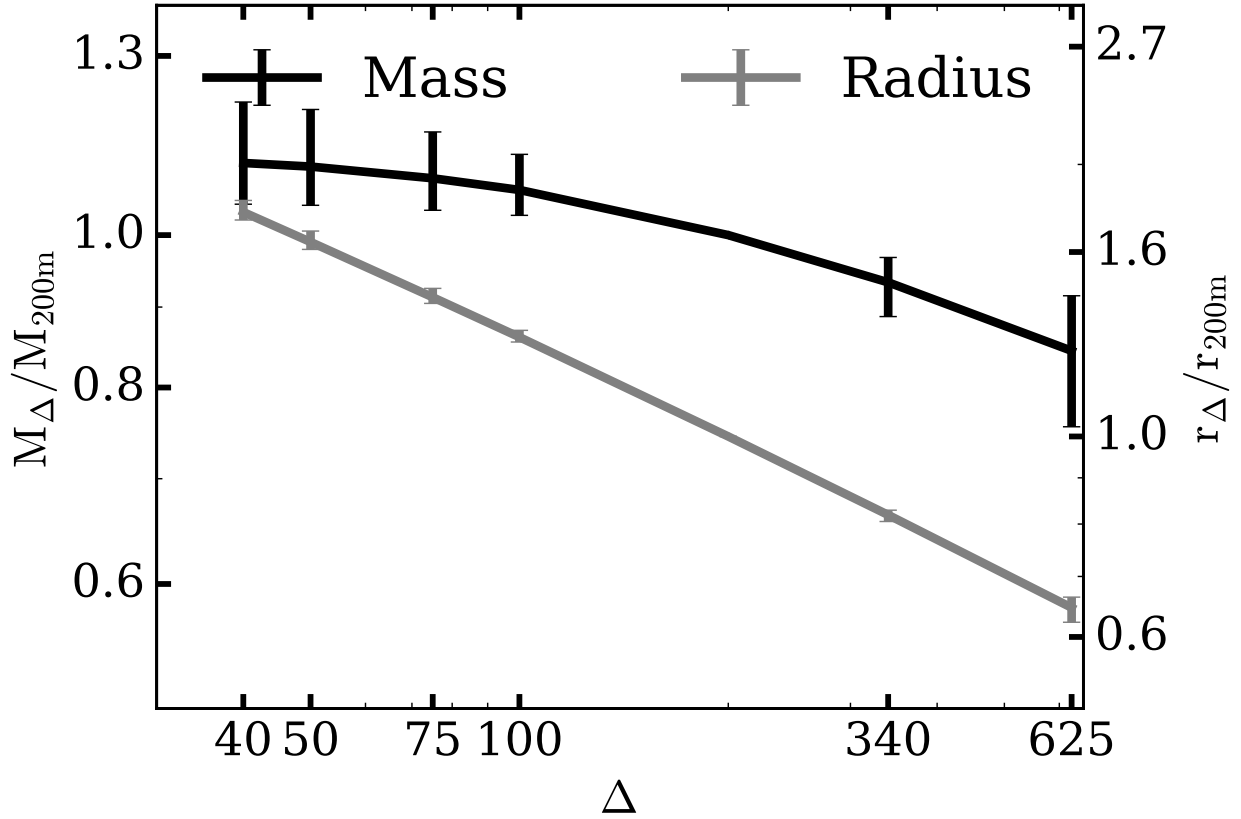


Figure 5 The ratio of halo properties as a function of  $\Delta$  in the L0250 catalog. The sample contains all host halos with masses greater than  $M_{200m} \geq 7.1 \times 10^{11}$ . The black (dark gray) line shows the median value of the ratio of the halo mass (halo radius) at a value of  $\Delta$  to the value at  $\Delta = 200m$ . The error bars contain 68% of values of this ratio for the sample. These ratios are very mild functions of mass, but all masses are stacked in this plot.

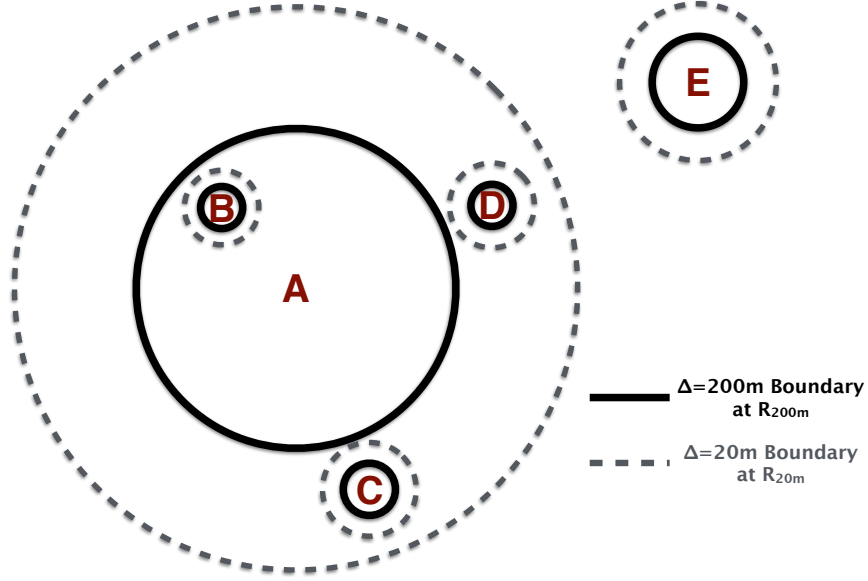


Figure 6 A qualitative illustration of the choice of halo definition. The figure shows five halos, labeled by the letters A-E. Halo A is the largest halo in the illustration. The solid halo boundaries correspond to the halo radii defined with respect to an overdensity of  $\Delta = 200\text{m}$ , namely  $R_{200\text{m}}$ . The dashed boundaries correspond to the halo radii defined with respect to an overdensity of  $\Delta = 20\text{m}$ ,  $R_{20\text{m}}$ . In all cases, halos become larger as  $\Delta$  decreases. Halos A and E are host halos according to both halo definitions. Halo B is a subhalo of halo A according to both halo definitions. Halos C and D are distinct host halos according to the  $\Delta = 200\text{m}$  halo definition, but they are reclassified as subhaloes of halo A according to the  $\Delta = 20\text{m}$  halo definition.

(the so-called “cusp-core” problem (de Blok, 2010, e.g.)), it serves as a useful tool for many semianalytic models. As such, halo concentration is defined with respect to the density profile of the halo,

$$\rho(r) = \frac{\rho_0}{\left(\frac{r}{r_s} + 1\right)^2}, \quad (3.2)$$

where  $\rho_0$  is a density scale and  $r_s$  is the scale radius, both of which vary from halo to halo. The latter parameter is of interest for halo concentration as it represents where the profile transitions from a steep “cuspy” profile to that of an isothermal sphere. Halo concentration is traditionally defined as

$$c_{\text{NFW}} = \frac{R_\Delta}{r_s}. \quad (3.3)$$

As such, this can be calculated with the inclusion of the scale radius,  $r_s$ , which is determined by ROCKSTAR by binning halo particles and carrying out a maximum-likelihood analysis to find a best fit NFW profile.

I further choose to use a “velocity-defined” halo concentration (Prada et al., 2012; Klypin et al., 2016),

$$c_V = \frac{V_{\text{max}}}{V_\Delta}, \quad (3.4)$$

where  $V_{\text{max}}$  is the maximum circular velocity of the halo and  $V_\Delta$  is the velocity of the halo at the halo radius,  $R_\Delta$ . Both of these values are calculated from the quantity  $\sqrt{GM(r)r^{-1}}$ ; the shallow dependence of this function on radius means there is minimal error due to finite sampling. This quantity,  $c_V$  has the advantage of being non-parametric and is more robust to halo density profile parameterization and fitting procedures. It is directly related to the NFW profile; if this profile is assumed, the halo concentrations can be related as:

$$c_V = 0.465 \left[ \frac{\ln(1 + c_{\text{NFW}})}{c_{\text{NFW}}} - \frac{1}{1 + c_{\text{NFW}}} \right]^{-1/2}. \quad (3.5)$$

For considering halo assembly bias, halo concentration is a useful parameter for many reasons. Halo concentration is known to be highly correlated with halo environment in standard halo definitions (Lee et al., 2017, e.g.). It is further of interest for modeling galaxy clustering and gravitational lensing statistics (and their cross-correlations). Galaxy clustering requires this indirectly as satellite galaxies in their groups may not necessarily trace the mass densities of their host halos. Gravitational lensing finds probing the underlying

mass density field as the primary goal and halo concentrations are thus of direct consequence for these predictions.

Additionally, halo concentration is known to be strongly correlated with the formation time of dark matter halos. Earlier forming halos trend toward higher values of halo concentration at fixed halo mass (Wechsler et al., 2002, 2006; Zhao et al., 2003, 2009). Exploring the impact of halo concentration allows us to probe the formation history of the halo within a single physical snapshot. The advantages to this computationally are significant; if one were to probe many different halo definitions, the procedure of calculating a halo merger tree for each halo definition is a not inconsiderable computational load. Utilizing halo concentration allows us to limit ourselves to snapshots at redshift  $z = 0$  and still explore a little of the halo’s formation history.

Note that while halo concentrations are known to correlate with halo formation time, there is significant scatter and this correlation may itself depend on halo environment. This limits our ability to draw conclusions on the environmental dependence of halo formation from our results on halo concentration. Exploration of halo formation time directly remains an intriguing topic for the future, especially in light of studies that demonstrate that assembly bias holds information in many different halo properties (Han et al., 2018).

A final motivation to explore halo concentration is that non-trivial manner with which halo concentration should inherently depend upon halo definition. Wechsler et al. (2002) demonstrated that clustering dependence on halo concentration changes sign near the collapse mass. I propose the following toy model by which halo definition should inherently result in changes to the both halo density profiles and the inferred halo concentrations. As subhalos preferentially reside on the outskirts of dark matter halos, their inclusion (or exclusion) has the potential to dramatically change inferred parameters. I demonstrate how this effect could take place in the Fig. 6 This might lead to halo concentration being the most sensitive parameter with respect to halo definition.

The upper-left panel of Fig. 7 shows the mean  $c_{\text{NFW}}-M_{\Delta}$  relation for halos with a fiducial  $\Delta = 200m$  definition. In each simulation, halos are considered only above a minimum mass threshold chosen to ensure that property measurements are not compromised due to insufficient mass resolution. The minimum mass threshold for each sample is depicted as

a downward arrow and are listed in Table 1 alongside the associated minimum number of particles. The upper-right panel of Fig. 7 shows the same relationship for the velocity-defined concentration,  $c_v$ . Halo concentration is a decreasing function of halo mass for each relation, consistent with the literature (Bullock et al., 2001; Macciò et al., 2007; Duffy et al., 2008; Prada et al., 2012; Klypin et al., 2016).

Fig. 8 demonstrates the strong correlation between NFW-defined and velocity-defined halo concentrations, with the prediction for a perfect connection assuming NFW halos as a red line. As is evident, these two parameters are strongly correlated and exhibit a  $\sim 6\%$  scatter around this correlation. This suggests that these two parameters roughly encode the same information about the underlying dark matter halo.

**3.2.1.4 Halo Shapes** The halo shape is another interesting halo parameter to study the impact of assembly bias. I study halo assembly bias with respect to the halo shape defined as,

$$s = c/a, \quad (3.6)$$

where  $c$  and  $a$  are the halo minor and major axis, respectively. These shapes are measured in ROCKSTAR using the method in Allgood et al. (2006), which calculates the modified inertia tensor,

$$M_{ij} = \frac{1}{N} \sum_{k=1}^N \frac{x_{i,k} x_{j,k}}{r_k^2}. \quad (3.7)$$

here  $x_{i,k}$  is the location of particle  $k$  along axis  $i$  with respect to the halo center and  $r_k$  is the distance between the particle and the halo center. The normalization factor of  $1/r_k^2$  is not part of the standard definition of the inertia tensor and serves the purpose of diminishing the influence of massive subhalos with large halocentric distances on the measure of halo shape. The calculation begins with all particles within a sphere of the halo radius and iterates the sample based on the previous shape measurement; particles associated with identified substructure are part of this shape measurement. The final inertia tensor is sorted for the eigenvalues which represent the squares of the principle ellipsoid axes with  $a > b > c$ .

The mean relations for halo shapes as a function of halo mass for  $\Delta = 200m$  are shown in the lower-left panel of Fig. 7. Note the mass thresholds (shown in Table 1) for halo

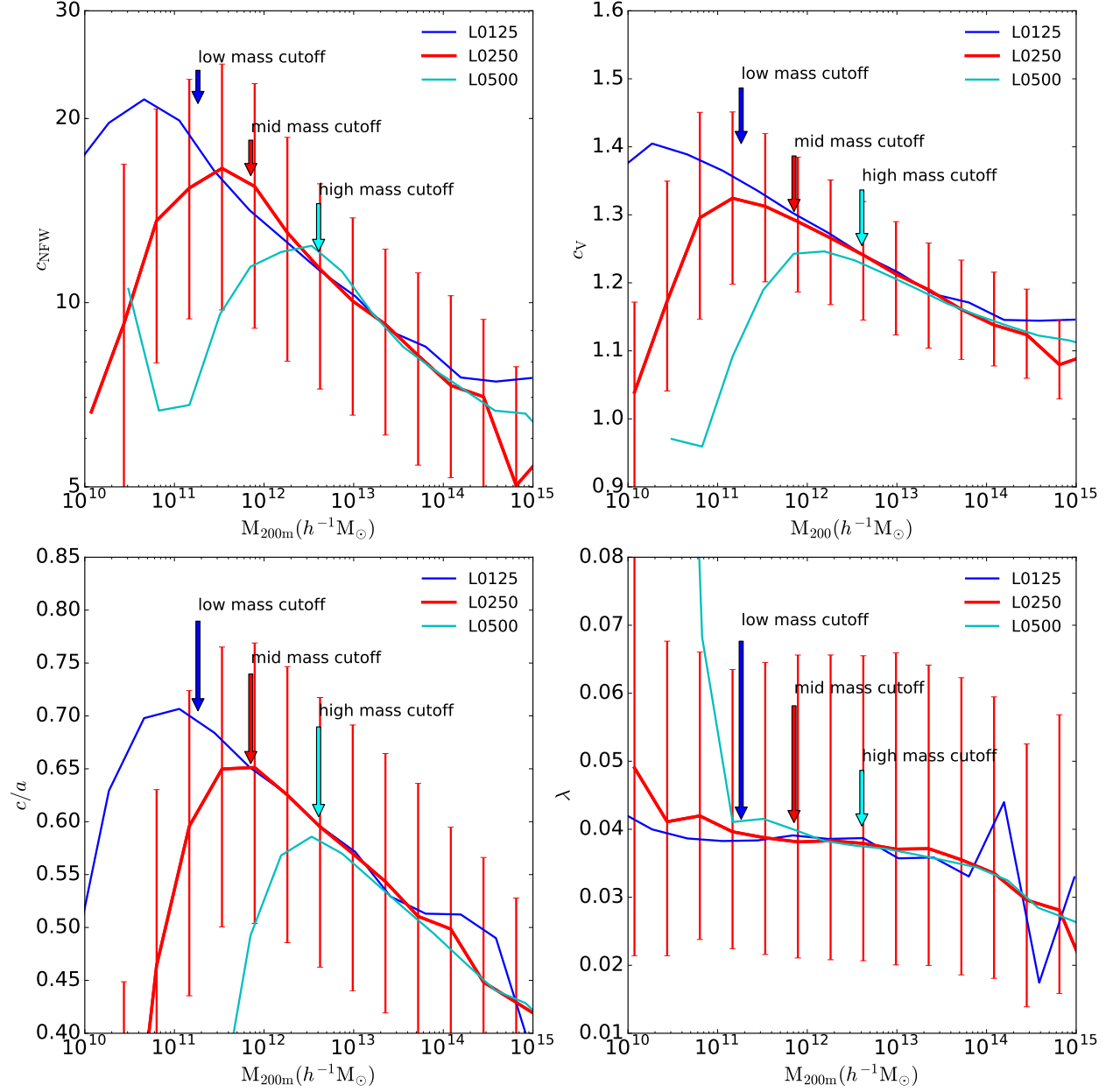


Figure 7 The relationship between mean halo properties and halo mass for each of our simulations with  $\Delta = 200m$ . These properties are the halo NFW-defined concentration, the velocity ratio defined concentration, the halo shape, and the halo spin. For example, in the top left panel, the blue (red/cyan) line corresponds to the concentration-mass relation from simulation L0125 (L0250/L0500). The red error bars show the 68% spread in parameter values within that mass bin for L0250. I show with arrows the minimum  $M_{200m}$  mass thresholds using the same color code as the concentration-mass relations.

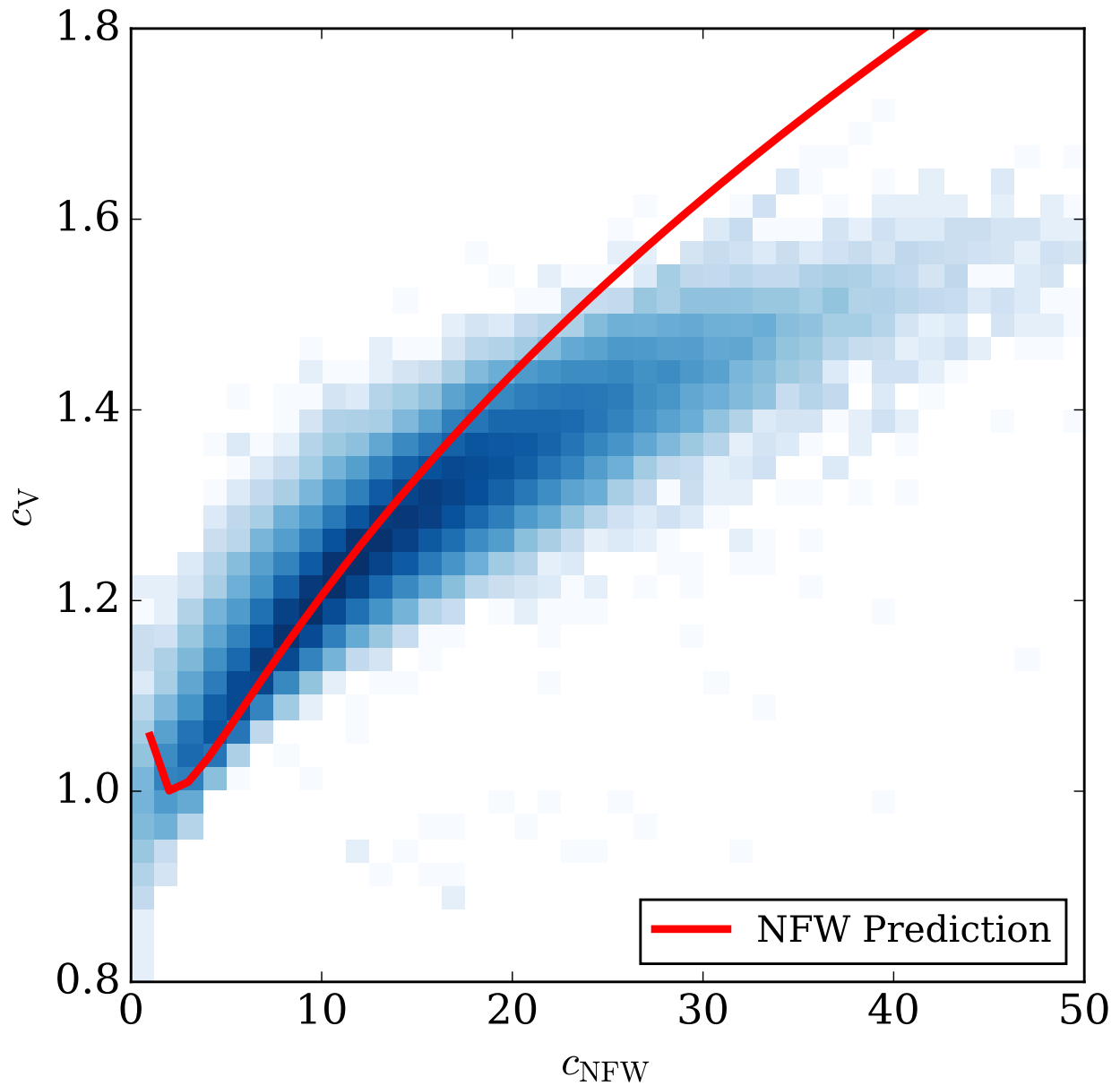


Figure 8 The relationship between the two different measures of concentration, using halos in L0250 defined with  $\Delta = 200m$ . The color scale, shown at the right, encodes the number of halos within a single two-dimensional bin in the  $c_{\text{NFW}} - c_V$  space. The dark (light) blue regions on the plot show where the most (fewest) halos exist with those values of the two concentration parameters. The white regions indicate where no halos hold these values. The scatter on this relationship ranges from 5% for intermediate concentration values, to a high of 13% at high masses. The red line shows the prediction from assuming an NFW profile.



shape and halo spin are considerably more aggressive than some in the literature; however, our summary statistics used coupled with the normalization discussed below minimize the impact of outliers. The relationship between halo shape and halo mass is such that halos become somewhat less spherical as halo mass increases and agrees with previous studies of halo shape (Jing & Suto, 2002; Allgood et al., 2006).

**3.2.1.5 Halo Spins** For understanding how halo angular momentum may be associated with halo clustering, I make use of the spin parameter,  $\lambda$ , as quantified by Peebles (1969),

$$\lambda = \frac{J\sqrt{|E|}}{GM_{\Delta}^{2.5}}, \quad (3.8)$$

where  $J$  is the the halo angular momentum,  $E$  is the *total* energy associated with the dark matter halo, and  $M_{\Delta}$  is the mass enclosed by the halo radius,  $R_{\Delta}$ . Essentially, this spin parameter measures the amount of energy in angular momentum in units of the amount necessary to support the halo against collapse through rotation. I use the measurements determined by ROCKSTAR for this calculation, which determines these three quantities within the halo radius with the bound particles of the halo. The mean relations of halo spin with halo mass can be seen in the lower-right panel of Fig. 7 along with the mass thresholds to avoid being compromised due to insufficient mass resolution (shown in Table 1). Note that halo spins are typically  $\lambda \sim 0.4$  and exhibit a very weak dependence with halo mass; this is consistent with the literature (Bullock et al., 2002; Macciò et al., 2007).

For much the same reasons as concentration dependent clustering, one could expect that the chosen halo definition has a large impact on the determination of halo spin. As seen in Fig. 6, it is quite possible that as size of a halo increases, a new subhalo can be subsumed into the larger structure. A new, high mass subhalo could potentially add a great deal to the angular momentum of a halo and increase the spin. The potential for this parameter to be sensitive to a choice of halo definitions motivates a thorough study of the clustering effects upon it.

Table 1 Minimum mass thresholds for each of our analyses. In the columns below each value of  $\Delta$ , I show the minimum host halo masses considered in units of  $h^{-1}M_{\odot}$ , the associated minimum number of particles, and the total number of host halos identified between this lower mass threshold and an upper mass threshold to avoid overlapping mass bins.

Simulation		$\Delta = 625m$	$\Delta = 340m$	$\Delta = 200m$	$\Delta = 100m$	$\Delta = 75m$	$\Delta = 50m$	$\Delta = 20m$
L0125	mass cut	$1.34 \times 10^{11}$	$1.67 \times 10^{11}$	$1.83 \times 10^{11}$	$1.94 \times 10^{11}$	$1.97 \times 10^{11}$	$2 \times 10^{11}$	$2.03 \times 10^{11}$
	# particles	837	1043	1143	1212	1231	1250	1268
	# halos	33623	29572	28467	27938	27754	27454	25329
L0250	mass cut	$5.23 \times 10^{11}$	$6.49 \times 10^{11}$	$7.1 \times 10^{11}$	$7.55 \times 10^{11}$	$7.66 \times 10^{11}$	$7.77 \times 10^{11}$	N/A
	# particles	402	499	546	580	589	597	
	# halos	88045	79795	78465	78851	79344	79152	
L0500	mass cut	$2.99 \times 10^{12}$	$3.71 \times 10^{12}$	$4.06 \times 10^{12}$	$4.31 \times 10^{12}$	$4.38 \times 10^{12}$	$4.44 \times 10^{12}$	N/A
	# particles	299	371	406	431	438	444	
	# halos	168742	158512	161339	169361	172175	175791	

### 3.2.2 Halo Samples

In practice, the mean relations between the various halo properties and the mass thresholds for these analyses must be determined separately for each combination of simulation, halo property (e.g.,  $c_{\text{NFW}}$ ,  $c_V$ ,  $\lambda$ , or  $s$ ), and halo definition (i.e., value of  $\Delta$ ). For each analysis, I set mass thresholds in order to avoid the regime in which halo parameters are not well measured due to resolution limits of the simulations; I draw attention to the deviation in the upper-left panel of Fig. 7 as an example of this. The NFW defined halo concentration follows an approximate power law with halo mass; however, at low particle numbers it is difficult to fit to an NFW profile and significant deviation from the mean relation exists. Simulations with larger number of particles at lower masses do not demonstrate this same effect, implying that this is primarily driven by simulation resolution.

For ease of comparison between halo definitions, I choose to use a single mass threshold for each simulation and for each value of  $\Delta$ . The mass thresholds are chosen to simultaneously minimize resolution effects and to include a similar population of halos above the threshold for each value of  $\Delta$ . I summarize the mass thresholds I have used for a subset of  $\Delta$  values in Table 1. Note that because halo abundance is a rapidly-declining function

of halo mass, the statistics quoted are always dominated by halos near the lower edge of the mass range. All statistics have been recomputed with mass threshold samples (rather than mass bins) and I find results that are both qualitatively and quantitatively similar. At most values of  $\Delta$ , the minimum mass thresholds are driven by the requirement that the halo properties do not suffer significantly from finite resolution effects. Be warned to the fact that the mass of an individual halo will vary as  $\Delta$  varies. This effect can be seen in Fig. 5, which demonstrates that while a decreased value of  $\Delta$  leads to larger masses on average, there is a scatter due to changes in halo identification. Roughly speaking, the threshold masses in Table 1 vary in such a way that the same physical objects are selected at each halo definition.

**3.2.2.1 Property Normalization** As seen in Fig. 7, many of these secondary halo properties exhibit strong mass dependence. Any analysis of clustering needs to be able to separate out this trend in order to avoid capturing the well known effect that high mass halos cluster stronger than low mass halos. In order to avoid accidentally drawing inference from this effect, I subtract off this gross mass dependence from each property with the following procedure:

1. Divide all halos above the mass threshold for resolution into equally spaced bins in logarithmic mass.
2. Calculate the rank of each halo with respect to an auxiliary halo property, from 1 to  $N$ , where  $N$  is the total number of halos in the mass bin.
3. Normalize each rank by the total number of halos,  $N$ .

In this manner I replace each halo property value with a percentile ranking at fixed mass. A halo with a normalized concentration of 0.74 has a concentration higher than 74% of halos at fixed mass. By construction, this removes the underlying mass trends in the initial data, allowing me to narrow in on the impact due to the halo properties themselves. This method is both qualitatively and quantitatively robust to different choices of bin widths (e.g. more bins, necessitating smaller bins) and binning schemes (e.g. equally populated bins), making little changes in my final results. The new formalism of using normalized ranks between 0 and 1 also provides computational benefits to be discussed later.

### 3.3 MEASURES OF ASSEMBLY BIAS

Halo assembly bias is defined as an excess clustering from that of the entire sample when subsampling along a secondary property. Measuring this can be done in several ways, of which I look at two that are fairly common in the literature. The first is that of using the correlation function directly; from this, a relative bias can be constructed. This method does require several fairly arbitrary choices, however, which leads to potential conclusion in the literature when comparing to other results. To evade this problem, I further make use of marked correlation functions; simply, these are correlation functions that have been weighted by an additional property. I discuss both in detail below, as well as how to identify assembly bias using these methodologies and `halotools`, a Python-based code that I use to carry out the numerical calculations.

#### 3.3.1 Correlation Function

The standard correlation function<sup>1</sup> (CF) used in the field of cosmology measures the probability that two objects are separated by a given distance. In this specific case, the objects that I am choosing to focus on is that of the dark matter halo. The calculation is carried out using the `halotools` function with an estimator of

$$1 + \xi(r) \equiv \text{DD}(r)/\text{RR}(r), \quad (3.9)$$

where  $\text{DD}(r)$  is the number of halo pairs with a given separation,  $r$ , and  $\text{RR}(r)$  is the number of randoms with the same separation. The random positions are drawn from a uniform distribution and has been chosen to have the same number of randoms as the examined halo sample. I specifically examine the range of  $r = 3$  to  $10h^{-1}\text{Mpc}$ , matching the range at which the impact of halo assembly bias is typically observed. The impact of halo assembly bias can then be determined by determining if different sampling cuts for a given property,  $p$ , lead to different values of clustering. At the simplest level, we choose to examine how behavior of

$$\frac{\xi_{p,80\%} - \xi_{p,20\%}}{\xi_{\text{all}}}, \quad (3.10)$$

---

<sup>1</sup>Formally a 2-point autocorrelation function.

where  $\xi_{p,80\%}$  is the clustering of the 80<sup>th</sup> percentile of halos in property  $p$  (the 20% with the highest rankings) and  $\xi_{p,20\%}$  is the clustering of the 20<sup>th</sup> percentile of halos in the same property (the 20% with the lowest rankings).

In order to determine the impact of halo assembly bias for a wide variety of halo definitions, I choose to examine the relative halo clustering bias of high- $p$  halos, where  $p$  is the property of interest, to all halos:

$$b^2(r) = \xi_{p,80\%}/\xi_{\text{all}}, \quad (3.11)$$

where  $\xi_{p,80\%}$  is again the clustering of the 80<sup>th</sup> percentile of halos in property  $p$ . Choosing different percentile values for the cut does quantitatively change this measure, but does not qualitatively change the results. Errors are then estimated on the relative bias by recomputing  $\xi_{\text{all}}$  using randomly-selected subsamples of equal number to the subsample used to compute  $\xi_{p,80\%}$ , that is, using one fifth of all halos.

### 3.3.2 Marked Correlation Function

In addition to the more traditional correlation function, I utilize a Marked Correlation Function (MCF) to examine auxiliary halo properties. MCFs quantify the manner in which a halo property (the “mark”) correlates among halo pairs as a function of a given separation between the pairs. MCFs have the advantage that they effectively stack signal from all values of the halo auxiliary property, or mark, in contrast to selecting subsets of haloes based on the auxiliary property. This removes one of the (somewhat) arbitrary choices discussed in the previous section on CFs. MCFs also stack signal from all environments and do not require any specific definition of halo environment in order to detect “environmental” trends that are usually referred to as assembly bias in the literature. MCFs have been used in many previous papers to quantify environmental dependence of halo properties other than mass (Beisbart & Kerscher, 2000; Faltenbacher et al., 2002; Sheth & Tormen, 2004; Sheth, 2005; Skibba et al., 2006; Harker et al., 2006; Wechsler et al., 2006; Mao et al., 2015a).

For a specific halo property, or mark  $m$ , I use the MCF normalization of Wechsler et al. (2006), namely

$$\mathcal{M}_m(r) \equiv \frac{\langle m_i m_j \rangle_{i,j \in P(r)} - \langle m \rangle^2}{\text{Var}(m)}, \quad (3.12)$$

where  $\langle m_i m_j \rangle_{i,j \in P(r)}$  is the mean of the product of two marks of a pair of halos separated by a distance about  $r$ ,  $\langle m \rangle$  is the mean of the mark for all halos, and  $\text{Var}(m)$  the variance of the mark for all halos.

In the absence of any correlation between a halo property among neighbors of a separation  $r$  away,  $\mathcal{M}_m(r) \simeq 0$  and its absolute value would be much less than 1. Deviations of the MCF from zero indicate such correlations exist and the magnitude of  $\mathcal{M}_m(r)$  gives the excess of the mark among pairs compared to the one-point mean of the mark  $\langle m \rangle$  in units of the one-point variance. The marks that I use are the normalized ranks of halo auxiliary properties described in Section 3.2.1, which are uniformly distributed between  $\frac{1}{N}$  and 1.

Identifying halo assembly bias using MCFs is fairly intuitive; absent halo assembly bias, the halo marks are expected to be uncorrelated across pairs. Additionally, as I make use of a normalized rank as the property mark, the error can be calculated using a single set of randoms. I assign randoms drawn uniformly between  $\frac{1}{N}$  and 1 to each halo and calculate the MCF for these marks. I then shuffle the position of the marks around halos to simulate the finite sampling impact, for 200 total realizations, allowing us to gain a strong understanding of the fluctuations associated with this finite sampling. By examining the 2% and 98% percentile limits of this sampling, I approximate a  $2\sigma$  error region that would be consistent with no halo assembly bias. While the results from MCF analysis need not necessarily agree with those of the CFs, I find that the two are in broad qualitative agreement for this analysis.

In the next chapter I will make use of these two analysis techniques on the [Diemer & Kravtsov \(2015\)](#) simulation data. I will demonstrate that each sample has assembly bias consistent with previous results with the literature and that both CFs and MCFs are valuable means of analyzing this. I will show that assembly bias is a strong function of halo mass, even with the removal of the underlying mass-secondary property relations.

## 4.0 HALO ASSEMBLY BIAS IN DIEMER ET AL.

The first focus of this work is to determine the strength of halo assembly bias within the [Diemer & Kravtsov \(2015\)](#) simulations that have been provided to us. This analysis can be done with both correlation functions and marked correlation functions; I choose to examine assembly bias driven by halo concentration with correlation functions in order to draw connections with similar works in the literature. Beyond that, I choose to focus on the marked correlation function, which is more robust to assumptions (e.g. no particular choice has to be made on where percentile cuts are made on a property). I focus on a fiducial definition that is common across many halo models and simulation analysis of  $\Delta = 200m$ ; this number has presumably become popular in the literature due to being a little more generous than the virial radius for spherical collapse in a  $\Lambda$ CDM universe and bearing similarities to the number examined for an Einstein-de Sitter universe.

### 4.1 CORRELATION FUNCTION RESULTS

The difference between the correlation function for the top 20% and bottom 20% of halos as ranked by NFW concentration for a fiducial  $\Delta = 200m$  halo definition for three simulation boxes (L0125, L0250, and L0500) can be seen in Fig. 9. Here the blue (magenta/cyan) solid line corresponds with the L0125 (L0250/L0500) simulation volumes. Accordingly, the blue line corresponds with the least massive halos ( $M_{200m} > 1.8 \times 10^{11}$ ) and the cyan line contains the most massive halos ( $M_{200m} > 4 \times 10^{11}$ ). In order to scale out the gross scale dependence of the CF, these differences are normalized by the CF of the clustering strength of the entire halo sample. The shaded band contains 98% of 200 CF ratios computed by randomly

subsampling halos in the L0500 sample and is a representative constraint on assembly bias non-detection.

If the clustering strength of halos was independent of NFW-defined halo concentration, all three lines in Fig. 9 would be clustered around zero with some statistical fluctuation (which is represented by the shaded band). Note that halos of small mass exhibit a considerably stronger degree of halo assembly bias than halos of larger mass, even after the removal of the underlying halo mass-concentration relation. This is consistent with concentration driven assembly bias as seen in the literature as a function of halo mass (Wechsler et al., 2006). This result is further robust to the choice of where to slice samples with respect to halo percentiles and is qualitatively similar to the result when examining the halo concentration as defined by the velocity ratio. These results are omitted for the sake of brevity, as they merely duplicate the results that will be examined more thoroughly with MCFs, which have distinct advantages for reproducibility.

## 4.2 MARKED CORRELATION FUNCTION RESULTS

### 4.2.1 Halo Concentration

The NFW concentration,  $c_{\text{NFW}}$ , MCF is shown in Fig. 10. The shaded band delineates the fluctuations expected from finite sampling as discussed in Section 3.3.2 and is generated for the L0500 simulation volume; this choice is representative of the fluctuations for the other simulation boxes as well. Qualitatively, Fig. 10 contains the same information that can be seen in Fig. 9. The least massive halos in the sample exhibit a stronger assembly bias with respect to halo concentration in comparison, while the most massive halos in the sample demonstrate a weaker assembly bias signal. It can be assumed that if a simulation with larger volume were available (containing more of the most massive halos), it would continue the trend of reverse halo assembly bias signal along these same trends. This is consistent with existing work in the literature (Wechsler et al., 2006; Sunayama et al., 2016).

The velocity concentration,  $c_v$ , MCF is shown in Fig. 11. The shaded band delineates



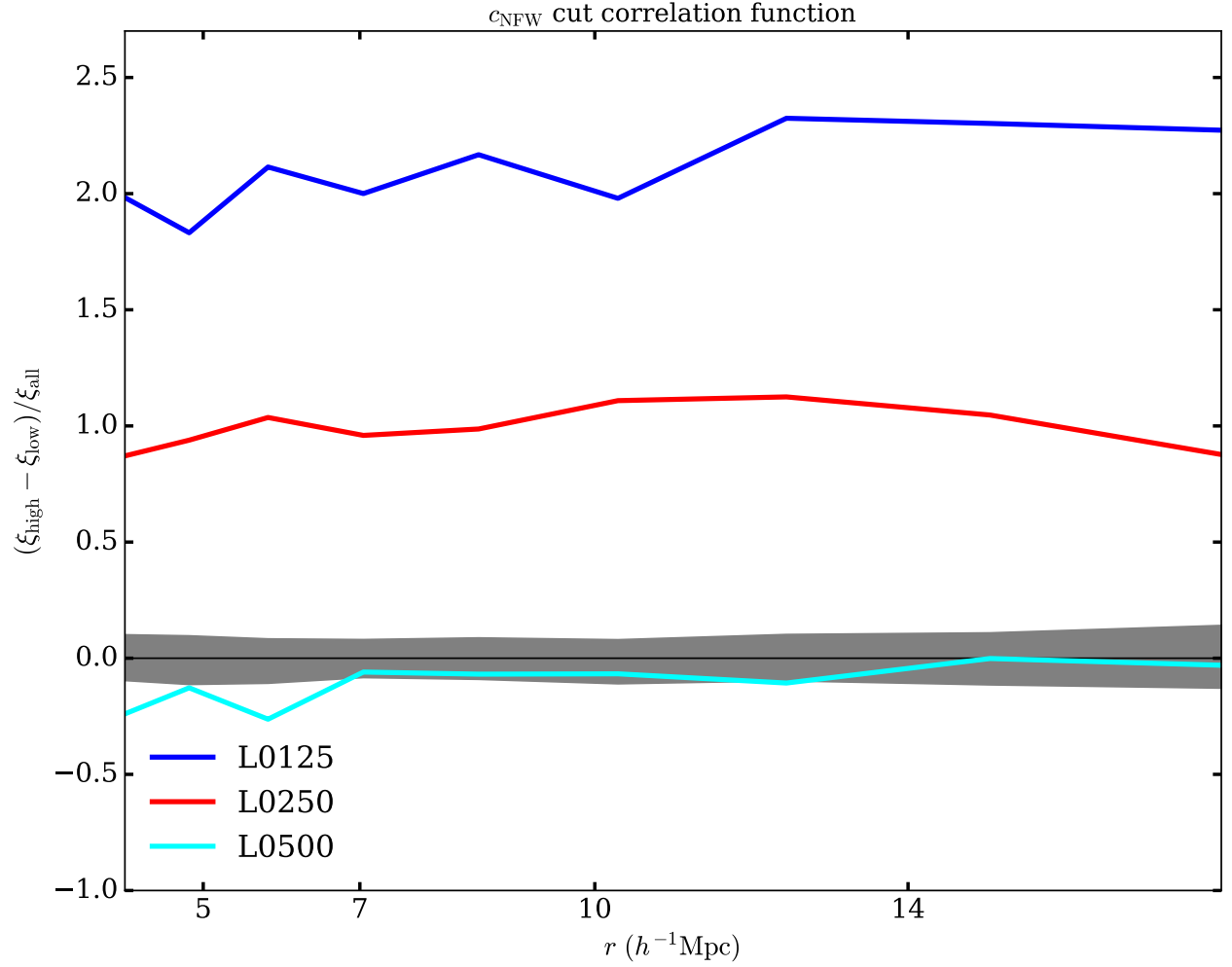


Figure 9 NFW concentration-dependent correlation functions. The solid lines plot the difference between the correlation function for the top 20% and the bottom 20% of halos by NFW concentration and a fiducial  $\Delta = 200m$  halo definition, normalized by the correlation function of the entire halo sample. The lines correspond to different simulation samples and accordingly different halo mass ranges, as seen in Table 1, with blue corresponding to the lowest mass halos and cyan corresponding to the highest mass halos. The shaded band contain 98% of 200 CF ratios computed from randomly subsampling halos in the L0500 sample and is a representative constraint on assembly bias non-detection.

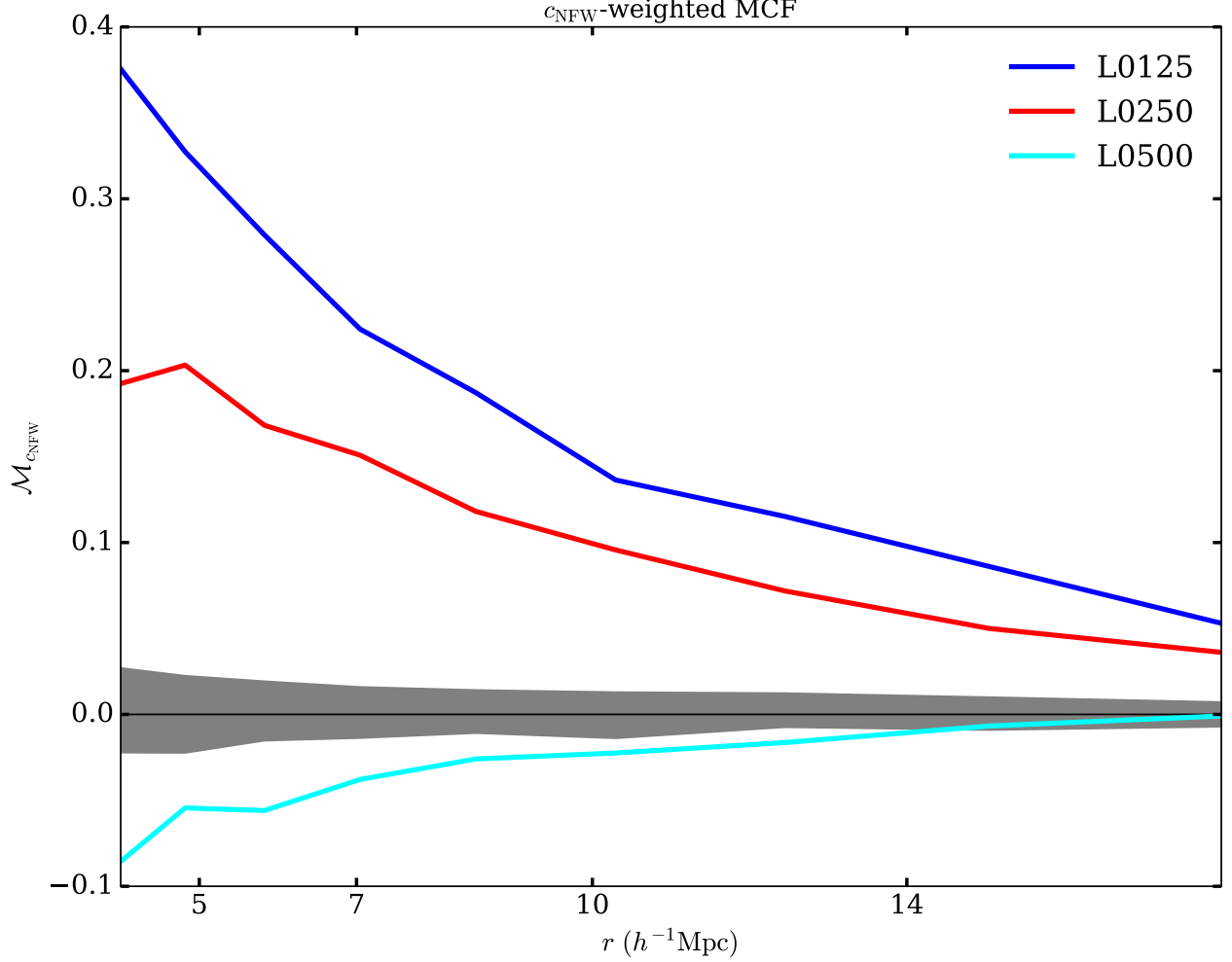


Figure 10 NFW concentration marked correlation functions. The solid lines plot the marked correlation function of halos weighted by NFW concentration for a fiducial  $\Delta = 200m$  halo definition. The lines correspond to different simulation samples and accordingly different halo mass ranges, as seen in Table 1, with blue corresponding to the lowest mass halos and cyan corresponding to the highest mass halos. The shaded band contain 98% of 200 MCFs computed from shuffling uniform random marks among the L0500 halo sample.

the same fluctuations expected from finite sampling as the previous Fig. 10 by construction. For ease of comparison, these two plots are plotted using the same range of possible MCF values. The reader should draw attention to the exceedingly minor differences between the two plots; the two marks have MCFs that are both qualitatively and quantitatively very similar. This should not be surprising, as  $c_{\text{NFW}}$  and  $c_{\text{V}}$  quantify largely redundant information about dark matter halos. This is suggestive that these results are not driven by intricacies in the fitting procedure to an NFW profile, which is known to be a poor description for halo profiles at large halocentric distances compared to  $R_{200\text{m}}$  (Diemer & Kravtsov, 2014).

In order to draw a physical intuition regarding these clustering result, consider that halo concentration is known to have a strong correlation with halo formation time. Halos with high values of concentration trend toward earlier formation times. A common result from hierarchical merger of dark matter halos is that overdense regions of the universe tend to evolve more rapidly than underdense regions. Thus, concentration driven assembly bias is a fairly natural expectation: those halos with high concentrations will tend to live in overdense regions and thus have enhanced clustering. My results demonstrate that this exists even when comparisons are drawn at fixed halo mass.

#### 4.2.2 Halo Shape

In addition to halo concentrations, I can explore the halo assembly bias with respect to less explored halo properties such as halo shape. The MCF results for this are demonstrated in Fig. 12, where the shaded band quantifies the fluctuations due to limited sample size. For all three simulations studied, the result agrees that the more spherical halos (and thus larger shape, as  $c/a = 1$  for a spherical halo) cluster together more strongly than less spherical halos. Further, this correlation does not have a strong dependence on halo mass, in contrast to the concentration driven assembly bias.

The physical picture for halo shape driven assembly bias is a little more difficult to motivate; one assumption that could be drawn is that those halos living in highly overdense regions are preferentially undergoing a larger number of mergers, which tends to create more

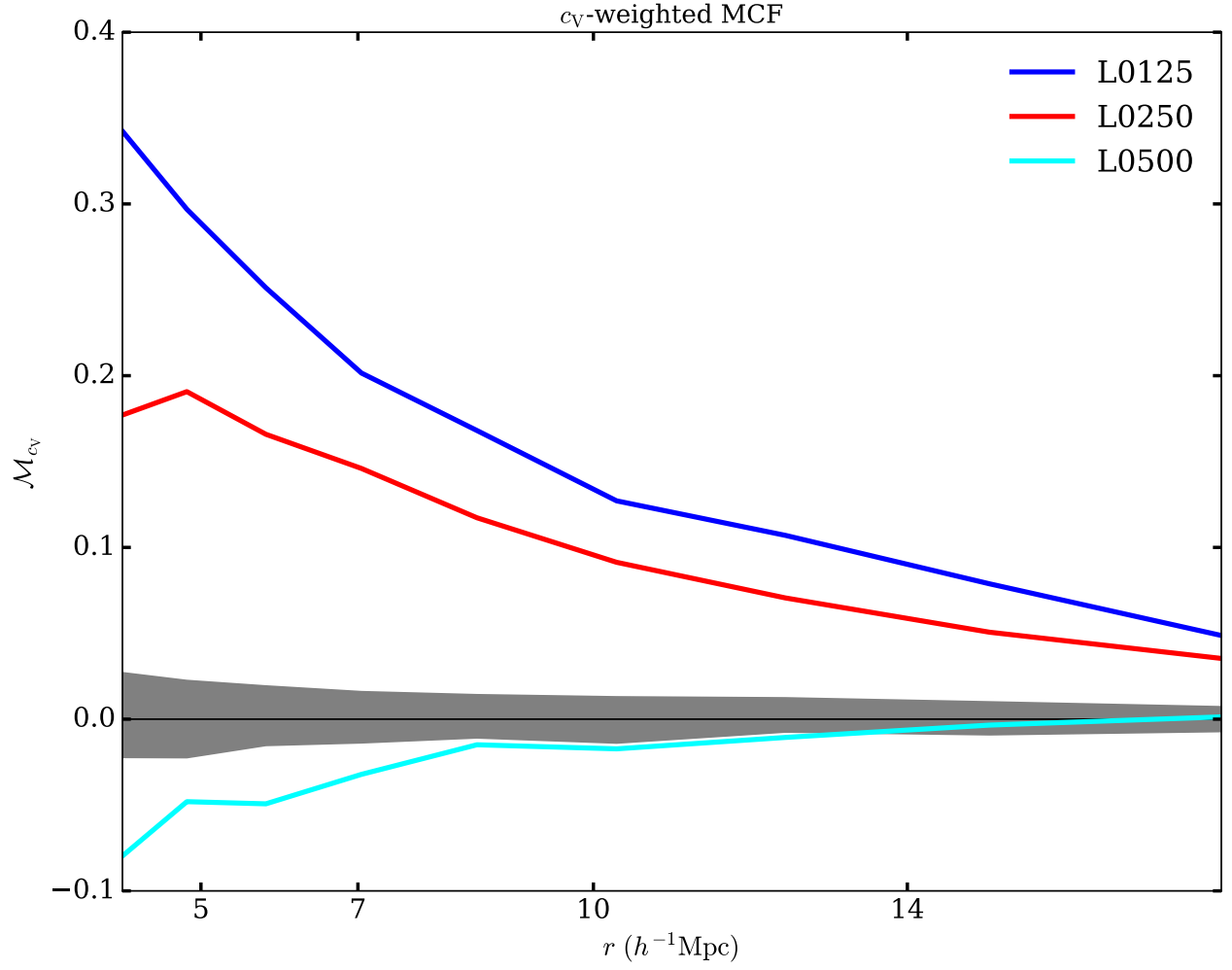


Figure 11 As Fig. 10, for the mark of halo  $c_V$ .

spherical halos over sufficient time. However, this is difficult to confirm without merger tree information which is beyond the scope of this research.

### 4.2.3 Halo Spin

Halo spin weighted MCFs are shown in Fig. 13. For the entire sample examined, halos with larger values of the spin parameter (that is, more energy in angular momentum compared to that necessary to support against collapse with rotation at fixed halo mass) correlate stronger than the zero assembly bias case. It is also evident in Fig. 13 that, contrary to the result of halo concentration, the clustering strength increases with respect to halo mass. This mass dependence is consistent with the previous literature on spin-dependent halo clustering, though the underlying physical reason remains a matter of intense debate.

Regarding the clustering itself, one might guess that those halos with more angular momentum preferentially live inside of overdense regions, such as inside the filamentary structure of dark matter distributions. As such, a preferred direction of infall for mergers may be able to add significant spin to dark matter halos. As with halo shape though, this is difficult to confirm without additional merger tree information.

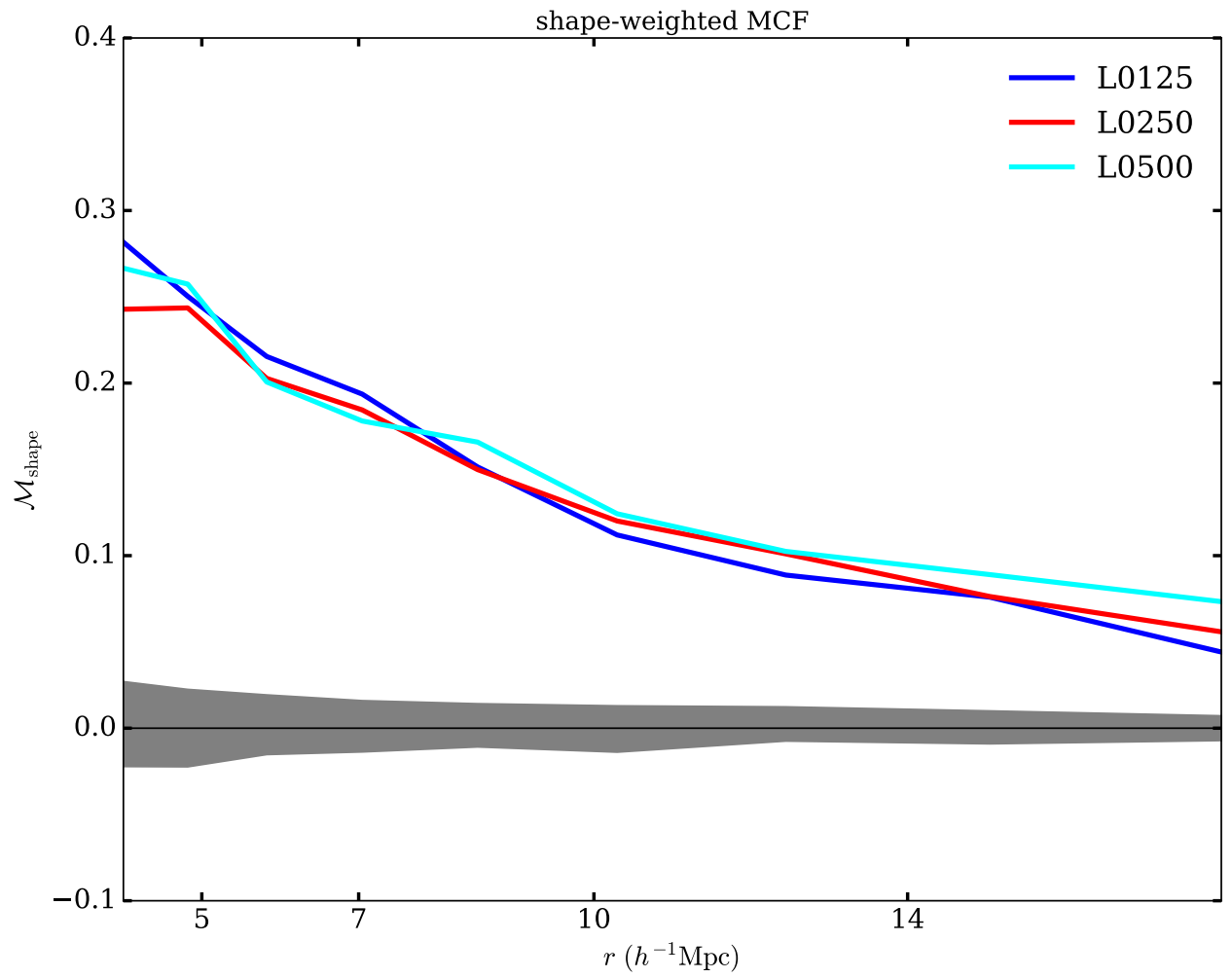


Figure 12 As Fig. 10, for the mark of halo shape.

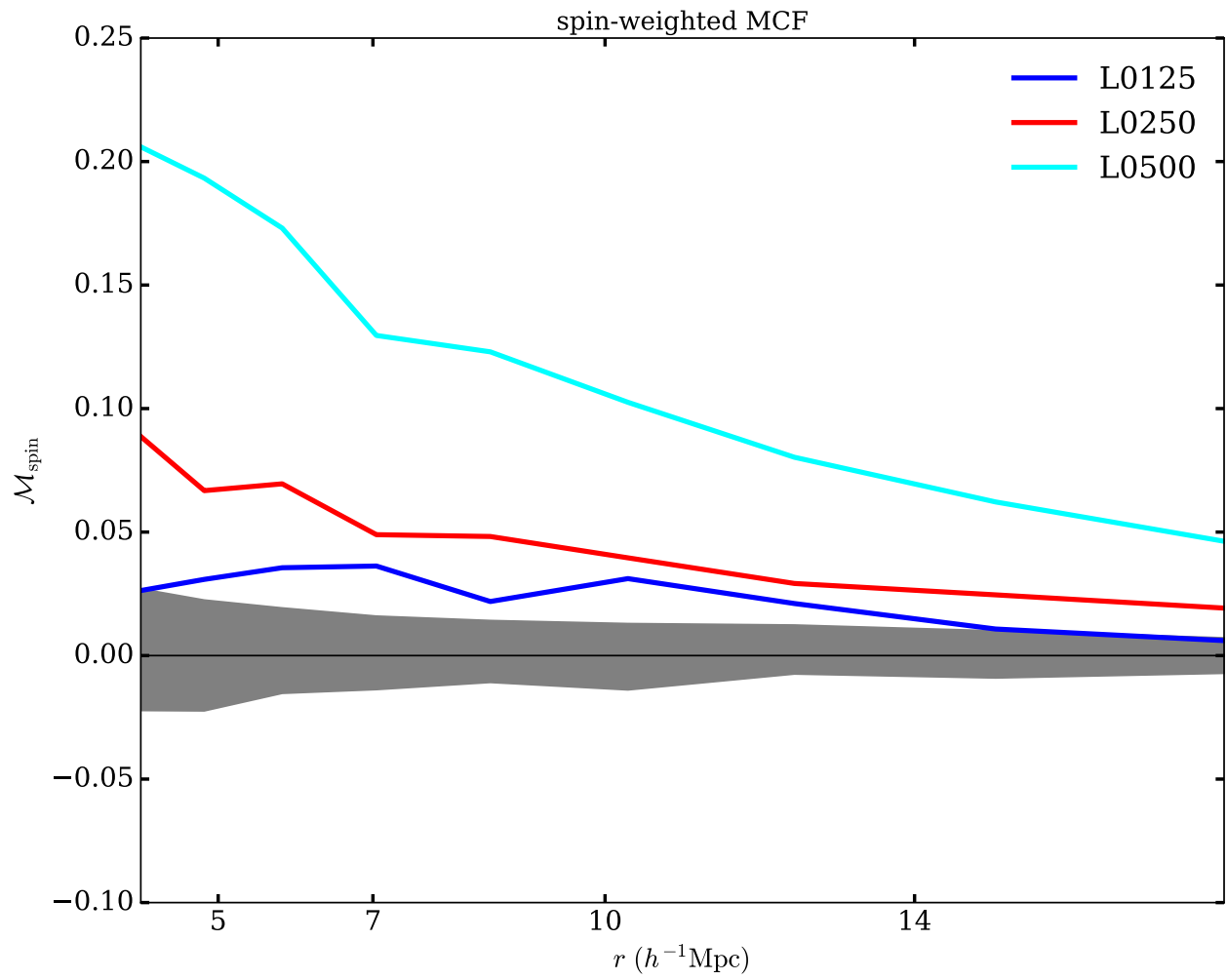


Figure 13 As Fig. 10, for the mark of halo spin.

## 5.0 HALO REDEFINITION IN DIEMER ET AL.

### 5.1 MOTIVATION

The previous Chapter 4 demonstrated the existence within the (Diemer & Kravtsov, 2015) simulations of halo assembly bias in agreement with the existing literature. With this in mind, I now set in mind to attempting to mitigate the impact of halo assembly bias through the use of a method referred to as “halo redefinition.” I direct the reader to Fig. 6, in which the size of the halo is changed with manipulation of the overdensity parameter,  $\Delta$ . A halo that increases in size may encompass nearby halos, leading them to be identified as substructure in a smaller value of  $\Delta$ .

The motivation behind this exploration is straightforward: the choice of halo definition as  $\Delta = 200m$  is not necessarily a physically motivated decision. It is well-known that not all gravitationally bound material associated with a halo lies inside of  $r_{200m}$  for a given halo. It is also well-known that not all halos identified are virialized. One can imagine a case where nearby, independent halos in  $\Delta = 200m$  are on their first infall into a larger halo and well on their way to becoming substructure; a larger halo size might encompass these physical effects and allow us to isolate halo assembly bias as being driven by this behavior, usually referred to broadly as “environmental effects”. Further, within the literature it is exceedingly common for the value of  $\Delta$  to change fairly arbitrarily; one prominent example of this is the choice of  $\Delta = 200c$ , where the critical density is used in place of the mean density. This roughly corresponds to  $\Delta = 625m$  at  $z = 0$ , which leads to a substantial difference in halo mass and size as seen in Fig. 5. With such arbitrary choices within the literature, it is worth an exploration for the sake of comparison alone!

Our methodology for this procedure is fairly straightforward. The ROCKSTAR halo



finder allows for different values of  $\Delta$  to be examined. I have re-run the halo finder to identify halos for an extremely broad range of halo definitions, from that of  $\Delta = 625m$  (which is comparable with using the critical density of the universe as opposed to the mean density) to  $\Delta = 20m$  (which is an exceedingly small value that has been unexplored in the literature). This necessitates a change in halo masses examined, which broadly matches the change of halo size and can be seen in Table 1.

## 5.2 RESULTS

### 5.2.1 Halo Concentration

I begin by studying the CFs of halos in our mass threshold samples, sub-selected by auxiliary properties. As an example, Fig. 14 exhibits the difference between the clustering strengths of halos in the top 20<sup>th</sup> percentile of NFW concentration compared to the halos in the bottom 20<sup>th</sup> percentile of NFW concentrations as a function of the overdensity parameter,  $\Delta$ , used to define the halos. Again, the gross scale dependence of the CFs have been normalized by the clustering strength of the entire halo sample. Now, the different colors express the different possible values of  $\Delta$ , with dark blue representing those halos with  $\Delta = 625m$  (a very small halo definition) and changing in color to light blue as  $\Delta = 50m$  (a very large halo definition). Note that the trends demonstrated in Fig. 9 can be seen now separated into three panels, allowing us to examine how halo redefinition changes observed halo assembly bias at fixed halo masses.

Draw your attention to the middle panel of Fig. 14. Note that the general trend of decreasing  $\Delta$  (and increasing halo size accordingly, as  $r_\Delta$  roughly varies in proportion with  $\Delta^{-1/3}$ ) is to reduce the strength of halo clustering in the high concentration sample compared to the low concentration sample. In particular, the red dashed line in this panel is the value calculated for a choice of halo definition of  $\Delta = 40m$  and has little detection of a clustering difference between the high and low concentration samples. It is clear from this behavior that both the strength and sense of halo assembly bias depend on the choice

of halo definition. While this may seem an obvious effect in hindsight, it is one that has not been well documented or studied in the existing literature, despite the clear impact of this choice.

Compare this across all three panels of Fig. 14. The rightmost panel (corresponding to the most massive halos) finds that a definition of  $\Delta = 250\text{m}$  has little detection of a difference between the two halo samples. In fact, reducing halo definition further than that starts to result in the low concentration sample exhibiting *stronger* halo clustering than the high concentration sample. The reverse behavior is true looking at the leftmost panel (and least massive halos), where a  $\Delta = 20\text{m}$  definition has very little detection of halo assembly bias. The reason for these changes is of interest and the implications are discussed below. Note that in Fig. 14 that the scale dependence persists regardless of choice of halo definition. The implication of this is that no single halo definition will mitigate halo assembly bias across the entire range of studied halo masses.

This same information is encoded into Fig. 15 and Fig. 16. As previously demonstrated, these two measures of halo concentration are tightly correlated and contain much of the same information about the underlying halo; as such, it is not surprising that their response to halo redefinition is near identical. Note again that in each panel, the scale dependence changes with both halo mass and halo definition, making it impossible for a single value of the halo definition to entirely remove assembly bias across the full sample of three simulation volumes.

### 5.2.2 Halo Shape

Fig. 17 demonstrates how halo redefinition impacts the measure of shape assembly bias. Note that the red dashed lines indicate the  $\Delta$  values that reduce assembly bias in the *concentration* dependent MCF; this has not been changed in these plots in order to give a comparison against how an optimal definition for one parameter may not be optimal for a secondary information; essentially, this is giving information that these additional parameters contain different information about the clustering of the dark matter halo. I draw the reader's attention to the impact of the choice of halo definition in Fig. 17. The

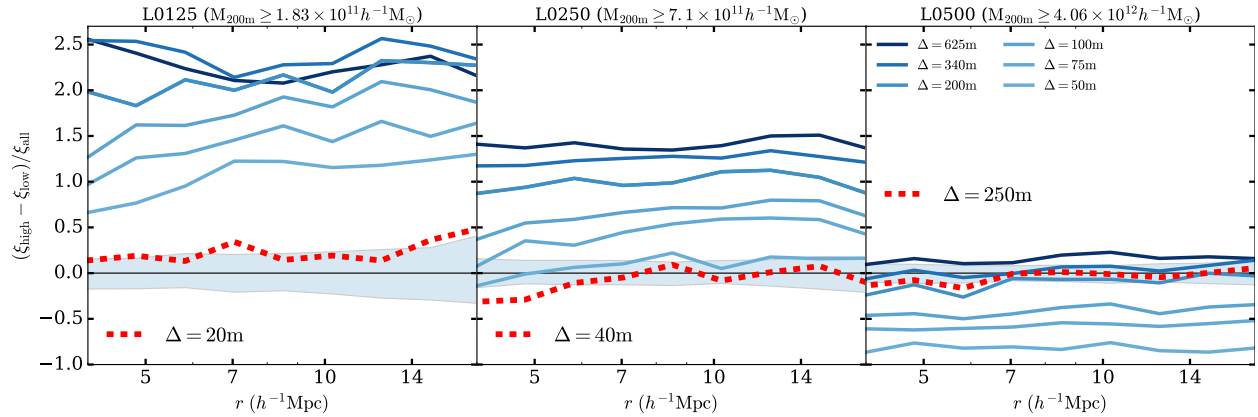


Figure 14 Concentration-dependent correlation functions. In each panel, the solid lines plot the difference between the correlation function for the top 20% and the bottom 20% of halos by NFW concentration, normalized by the correlation function of the entire halo sample. The lines correspond to different values of  $\Delta$ , with dark blue (light blue) corresponding to  $\Delta = 625\text{m}$  ( $\Delta = 50\text{m}$ ). The red dashed line in each panel corresponds to an overdensity that significantly reduces concentration-dependent halo clustering selected from the various values of halo overdensity that I explore. I will refer to these values throughout the text. The left (middle/right) panel shows the results for L0125 (L0250 /L0500) utilizing the low-mass (mid-mass/high-mass) halo threshold samples. To guide the reader, each panel is labeled by the minimum value of  $M_{200\text{m}}$  in the sample. Different values of  $\Delta$  correspond to different mass thresholds as shown in Table 1. The shaded bands in each panel represent the level of statistical fluctuations due to finite sampling. In particular, the shaded bands contain 98% of 200 CF ratios computed from randomly subsampling halos in the  $\Delta = 200\text{m}$  sample. In principle, each sample with a distinct  $\Delta$  should have a distinct error band, but in practice they are all very similar to that of the  $\Delta = 200\text{m}$  sample.

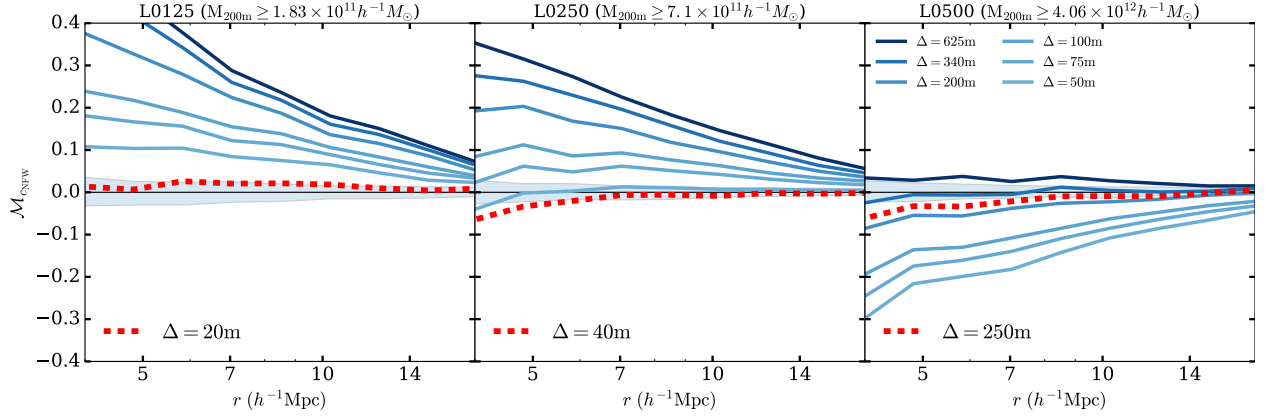


Figure 15 The marked correlation function for the concentration defined according to the NFW profile,  $c_{\text{NFW}}$ . The solid lines plot the marked correlation function using normalized ranks of NFW concentration as the mark. In each plot the lines correspond to different values of  $\Delta$ , with dark blue (light blue) corresponding to  $\Delta = 625\text{m}$  ( $\Delta = 50\text{m}$ ). The red dashed lines correspond to the overdensities that greatly mitigate assembly bias for concentration (the same values of  $\Delta$  as depicted by the red dashed lines in Fig. 14). The top (middle/bottom) panel shows the results for the L0125 (L0250/L0500) data set utilizing the low mass (mid mass/high mass) thresholds. To guide the reader, each panel is labeled by the minimum value of  $M_{200\text{m}}$  in the sample. Different values of  $\Delta$  correspond to different mass thresholds as shown in Table 1. The shaded bands in each panel represent the level of statistical fluctuations due to finite sampling. In particular, the shaded bands contain 98% of 200 MCFs computed from shuffling the halo marks among all of the halos in the  $\Delta = 200\text{m}$  sample. In principle, each sample with a distinct  $\Delta$  should have a distinct error band, but in practice they are all very similar to that of the  $\Delta = 200\text{m}$  sample.

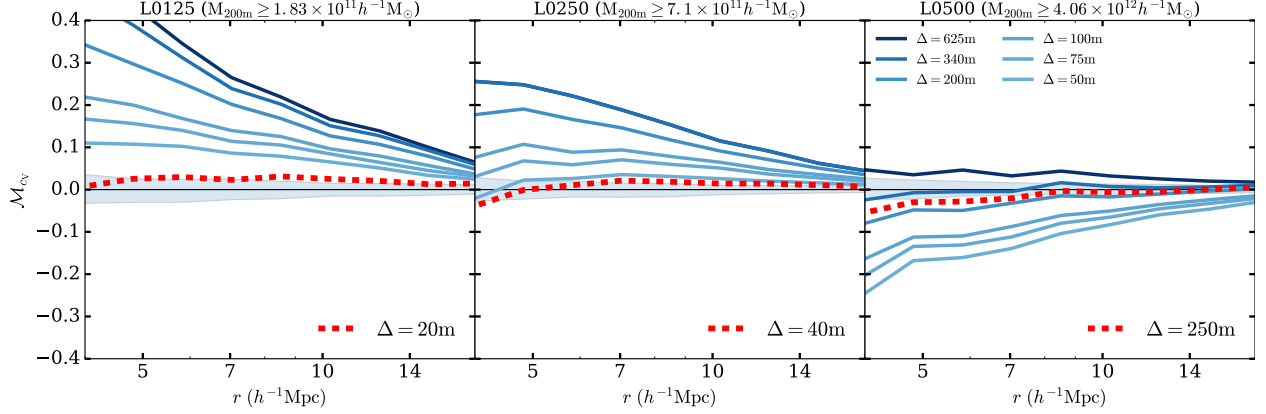


Figure 16 The same as Fig. 15 for the mark of  $c_V$ . Red dashed lines indicate the  $\Delta$  values that remove assembly bias from *concentration*.

smallest halo definition ( $\Delta = 625m$ ) demonstrates a larger assembly bias detection than the largest halo definitions, independent of mass bin. The direction of this effect matches the behavior of the halo concentration marks that were explored in Fig. 15 and Fig. 16. However, there exists no halo definition explored that removes the assembly bias impact on the scales that I study. While it is possible an even more extreme halo definition may move into that territory, the consequences of a halo with overdensity  $\Delta$  of that size is unexplored at this time.

### 5.2.3 Halo Spin

The results for halo redefinition on halo spin are shown in Fig. 18. First, note that the red dashed lines which correspond to removing halo assembly bias with respect to halo concentration are (again) not able to remove halo assembly bias with respect to halo spin. I draw reader attention to the reversal of the trend with halo redefinition compared to the other three explored halo properties. As I reduce the value of  $\Delta$  (therefore increasing the size of the identified dark matter halos), the enhanced clustering with respect to halo spin is *increased*. One speculative means of explaining this result is that material on the fringes of the halo can carry a large amount of angular momentum with respect to the halo center (due to the large value of the impact parameter to the halo center). As such, increasing the size of

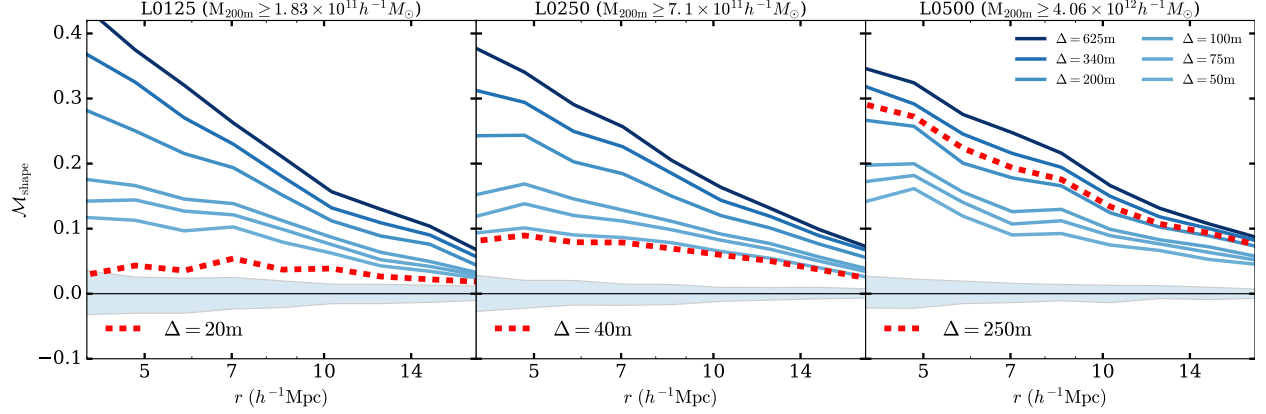


Figure 17 The same as Fig. 15 for a halo shape MCF. Red dashed lines indicate the  $\Delta$  values that remove assembly bias from *concentration*.

the halo may make the halo spin parameter *more* sensitive to overdense halo environments rather than weakening the environmental dependence. Follow-up work is necessary in order to verify whether this speculative theory is accurate and this will be explored in the future.

### 5.3 SUMMARY OF HALO CLUSTERING AND HALO REDEFINITION

#### 5.3.1 Halo Bias and Halo Redefinition

Various forms of assembly bias, and their mass dependences, have already been identified in the literature. This work confirms these previous results for halo concentrations (Wechsler et al., 2006; Faltenbacher & White, 2010; Sunayama et al., 2016), halo shapes (Bett et al., 2007; Hahn et al., 2007b,a; Faltenbacher & White, 2010; Lacerna & Padilla, 2012; van Daalen et al., 2012), and halo angular momenta (Bett et al., 2007; Hahn et al., 2007b,a; Lacerna & Padilla, 2012). It emphasizes the fact that auxiliary property dependent halo clustering, which I refer to as assembly bias, is strongly dependent upon halo definition. Fig. 19 summarizes these results for each of the properties of interest. For ease of comparison with other works in the literature, I now plot the relative halo bias of a high- $p$  halo sample to that of the overall halo sample, where  $p$  is the halo property of interest. As previously discussed,

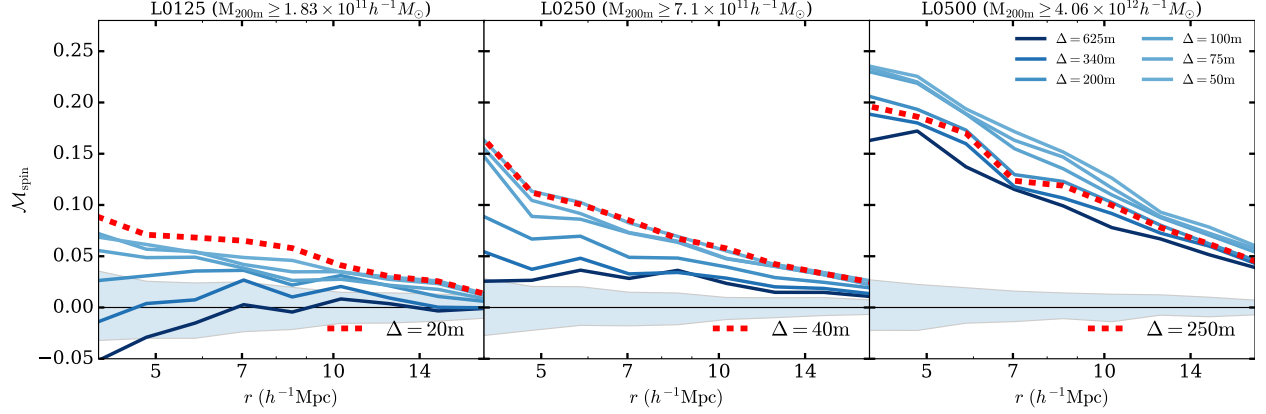


Figure 18 The same as Fig. 15 for a halo spin MCF. Red dashed lines indicate the  $\Delta$  values that remove assembly bias from *concentration*.

the high- $p$  sample consists of halos above the 80<sup>th</sup> percentile and above in a given property at fixed halo mass. Though quantitatively different, a different choice of percentile quantity does not qualitatively change the results. I estimate the errors on the relative bias by re-computing  $\xi_{\text{all}}$  using randomly-selected subsamples of equal number to the subsample used to compute  $\xi_{p,80\%}$ , that is, using one fifth of all halos. The error bars show the 68% range, centered on the median, about which the values of  $b^2(r)$  lie.

Each panel of Fig. 19 allows one to focus on the mass dependence of halo clustering and the dependence on halo definition. Focus first on the top two panels, which deal with our two halo concentration measurements, NFW concentration and velocity ratio concentration. Both measures of concentration exhibit a strong mass dependence to assembly bias. Halos at high masses ( $M_{\Delta} \geq 10^{13} h^{-1} M_{\odot}$ ) exhibit assembly bias that is quite distinct from that of low mass ( $M_{\Delta} \leq 10 \times 10^{12} h^{-1} M_{\odot}$ ) halos. At high masses, high-concentration halos are less strongly clustered than low-concentration halos. At low-masses the sense of assembly bias is reversed and high-concentration halos cluster more strongly. Notice also that at high masses, concentration-dependent halo clustering is only a modest function of halo mass.

Concentration dependent halo clustering has a significant dependence upon halo definition. In particular, the strength of clustering of the halos above the 80<sup>th</sup> percentile in concentration varies by as much as  $\sim 50\%$  among the overdensities that I have investigated. For all halo definitions, concentration-dependent clustering exhibits a similar trend with

halo mass. For no value of  $\Delta$  does the trend with mass become significantly less prominent, suggesting that a simple mass-independent re-definition of halo boundaries alone cannot eliminate halo assembly bias. Concentration-dependent clustering changes sense at a mass that varies by an order of magnitude, from  $\sim 3 \times 10^{12} h^{-1} M_{\odot}$  to  $\sim 3 \times 10^{13} h^{-1} M_{\odot}$ , as  $\Delta$  varies from  $\Delta = 50$  to  $\Delta = 625$ . These significant variations in the strength of concentration-dependent clustering as a function of halo definition suggest that significant care must be taken in comparisons of various results in the extant literature.

The bottom two panels of Fig. 19 demonstrate that auxiliary property dependent clustering can behave in a markedly different manner depending upon the halo property under consideration. These two panels show shape (left) and spin (right) dependent clustering as a function of halo mass and halo definition ( $\Delta$ ). Each demonstrates halo assembly bias that is only weakly dependent on halo mass in comparison to concentration-dependent clustering. In the lower-left panel, note that changing to smaller values of  $\Delta$  results in reduced assembly bias, though no definition explored was sufficient to remove assembly bias entirely.

It is worth emphasizing that the  $\Delta$  dependence of assembly bias in Fig. 19 is non-trivial. At first glance, the reader may be tempted to think that each of the lines in any individual panel of Fig. 19 could be made to overlap by plotting the bias with respect to a common mass scale for each halo, rather than for  $M_{\Delta}$  for each value of  $\Delta$ . In other words, the reader may be tempted to think that the only reason that the lines in Fig. 19 do not overlap is trivially due to the shift in halo masses caused different  $\Delta$  (e.g., Fig. 5). However, this is *not* the case. Consider the panels of Fig. 19 depicting halo concentration. Suppose that I had chosen to plot halo relative bias as a function of the independent variable  $M_{200m}$ , the  $\Delta = 200m$  mass of each halo. For any individual object, if  $\Delta' < \Delta$ , then  $M_{\Delta'} > M_{\Delta}$ . Consequently, for any individual halo  $M_{200m} > M_{625m}$  while  $M_{200m} < M_{50m}$ . Therefore, representing the relative halo bias on a common mass scale (e.g.,  $M_{200m}$ ) for all halos moves the curves in Fig. 19 *further away from one another*. The analogous statement is true for all panels of Fig. 19. The point is profound because it demonstrates that the  $\Delta$  dependence of assembly bias is *not* caused simply by the shift in mass scale, but rather by the selection of host halos to include in the halo sample.



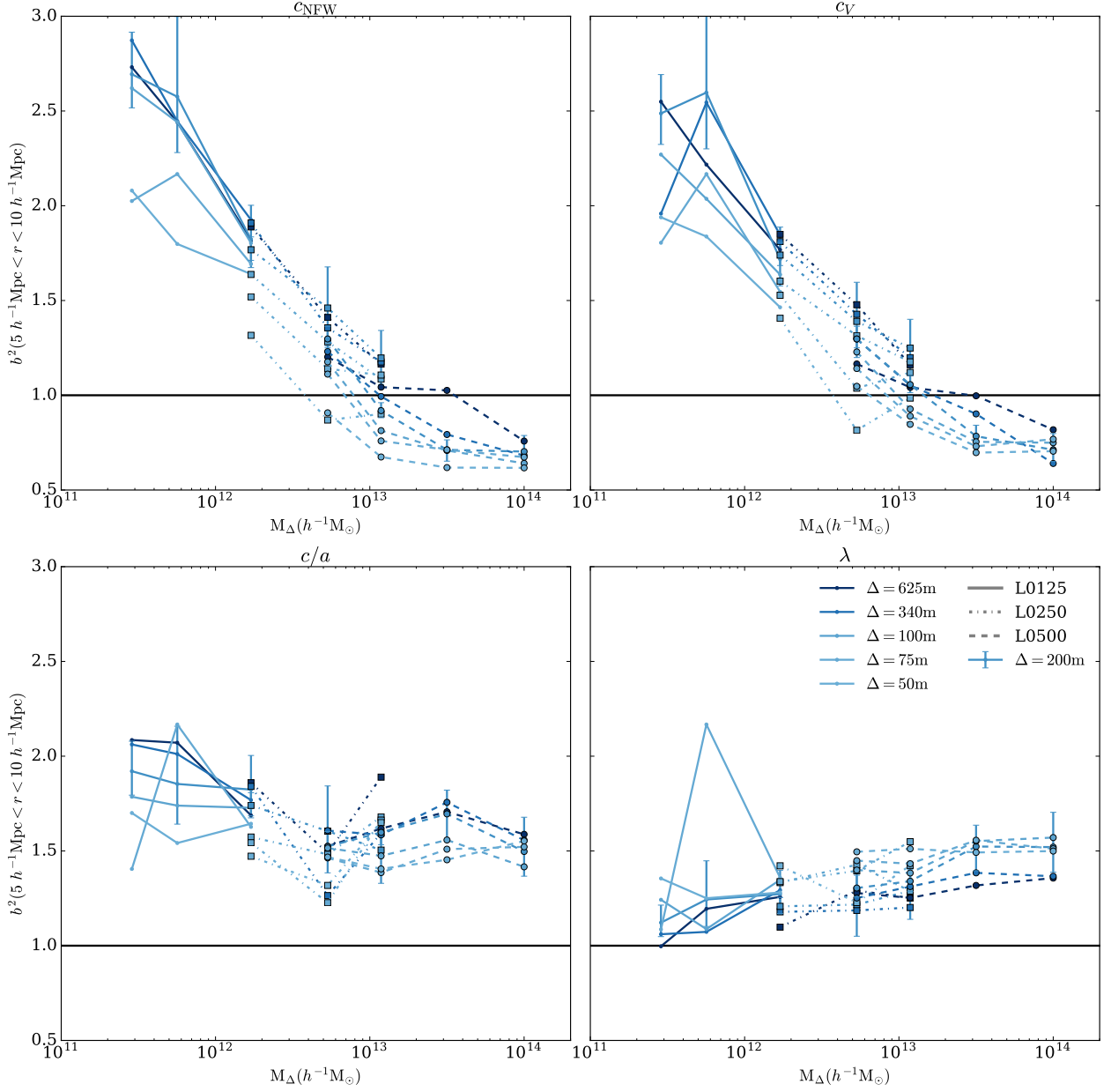


Figure 19 The relative clustering bias of halos as a function of halo mass for various halo auxiliary properties. I show the clustering strength of the highest 20-percentile according to each halo property relative to the population of all halos. Within each panel, I show clustering biases for six values of  $\Delta$ . The dark blue (light blue) line uses a halo definition drawn from  $\Delta = 625\text{m}$  ( $\Delta = 50\text{m}$ ). The solid (dot-dashed, dashed) lines use host halos from the L0125 (L0250, L0500) catalogs. The error bars on the  $\Delta = 200\text{m}$  samples are similar to the errors from other samples, which are not shown for clarity.

### 5.3.2 How Does Halo Redefinition Help?

Note that the results above indicate that one may mitigate some auxiliary property dependent clustering through a mass-dependent choice of halo definition. There are restrictions to this statement: notably, one cannot remove halo assembly bias across all halo properties. In this subsection I discuss the possible reasons behind the mitigation of concentration-dependent halo bias, as it is the most promising of the halo properties under this method.

Halo redefinitions may mitigate concentration-dependent halo bias for at least two reasons. The first reason is the physical motivation for exploring alternative halo definitions. In particular, it may be that alternative halo definitions (presumably with smaller  $\Delta$ ) provide a more effective grouping of objects that have been strongly affected by interactions into one single halo as illustrated in the cartoon in Fig. 6. If this is the case, alternative halo definitions may offer a more practical separation between the linear and highly nonlinear regimes and may represent a pragmatic step forward. However, alternative halo definitions may also reduce auxiliary property dependent clustering through a second mechanism. To be specific, it is possible that the details of measuring halo properties using these new halo definitions introduce new sources of noise into the measurements. These inherently noisier measurements then lead to reduced correlations between pairs. In this second case, the reduction in correlations simply arises because the property of interest may be *less* informative about the halo itself - a concerning proposition.

In the case of halo concentration, noise may be introduced in numerous ways. For example, the NFW concentration  $c_{\text{NFW}}$  is determined by a fit to the NFW profile. Inferred values of  $c_{\text{NFW}}$  will depend upon the degree to which the density profiles of the halos follow the NFW functional form within some radius  $R_\Delta$  that is different from traditional halo radii, such as  $\sim R_{200\text{m}}$ . At large halocentric distances ( $r \gtrsim R_{200\text{m}}$ ) halo profiles are known to deviate markedly from the NFW form. It may be possible to reduce assembly bias by redefinitions if one probes scales on which halos deviate from NFW in a way that is not well correlated with the interior structure of the halo (particularly the location of the NFW scale radius); however, such a reduction in assembly bias is of limited practicality because it arises from characterizing a halo by a quantity that is less informative about its interior structure.

Of course, it is worth noting that the velocity-defined concentration  $c_v$  is a non-parametric measure of concentration and should be less subject to such effects.

I explore in more detail the degree to which the mitigation of environmental effects by halo redefinition is due to the introduction of noise that is uncorrelated with environment into the measurement of halo properties. I describe one method for doing this in the remainder of this section. The cartoon of Fig. 6 may be a useful reference for the reader.

Consider first that all host halos that are present in the halo catalogs constructed from any one specific value of the overdensity threshold (e.g.,  $\Delta = 40m$ ) are also present as host halos in the halo catalogs constructed with all higher values of threshold density (e.g.,  $\Delta = 200m$  and, in fact, all  $\Delta \geq 40m$ ). The converse is not true because lowering  $\Delta$  increases halo radii so that host halos at higher values of  $\Delta$  may become subhalos at lower values of  $\Delta$ <sup>1</sup>. This reclassification as subhalos is the fate of halos C and D in Fig. 6 as the overdensity threshold is decreased from  $\Delta = 200m$  to  $\Delta = 20m$ .

Consider the exercise of matching halos across different catalogs, constructed using different values of  $\Delta$ . I match halos to a *baseline* catalog corresponding to a value of  $\Delta$  that best mitigates concentration-dependent clustering<sup>2</sup> at each mass bin. These baseline values of  $\Delta$  are those delineated by the dashed, red lines in Fig. 14–18. In this exercise, I consider the clustering of only those halo that are classified as host halos in the baseline catalog.

Specific examples may help to clarify this procedure. Consider, for example, the low-mass sample constructed from the L0125 simulation. The baseline overdensity in this case is  $\Delta = 20m$ . For the low-mass sample, I study the clustering of halos designated as host halos according to a  $\Delta = 20m$  halo definition. However, I assign these halos properties according to their definitions using other values of  $\Delta$ . For example, I may study the clustering of these halos as a function of the masses and concentrations defined using a  $\Delta = 200m$  halo definition. Referring back to Fig. 6, I would take the clustering of halos A and E using halo properties derived from the particles within  $R_{200m}$ . This is useful because it preserves the original halo properties defined using a conventional halo definition (such as  $\Delta = 200m$ ), but it removes those halos from consideration that I expect to be altered by interactions

---

<sup>1</sup>Indeed, this is largely the motivation for exploring various  $\Delta$ , as discussed previously.

<sup>2</sup>For my limited explorations of  $\Delta$ ; there may be a choice which better mitigates halo assembly bias.

with neighboring halos. In the cartoon of Fig. 6, this removes halos C and D and ceases to consider their impact on clustering. They are now subhalos in the  $\Delta = 20\text{m}$  halo catalog.

A second example may be useful and is relevant to the high-mass sample that I study, constructed from the L0500 simulation. In this case, the baseline overdensity is  $\Delta = 250\text{m}$ . I consider the clustering of all objects that are host halos in the  $\Delta = 250\text{m}$  halo catalog. If I define properties relative to any  $\Delta \geq 250\text{m}$ , then this is very much the same as in the previous example in that using the baseline catalog serves to eliminate some halos from consideration. However, for  $\Delta < 250\text{m}$ , some of the objects that I consider are subhalos. In the context of Fig. 6, this is analogous to studying the clustering of halos A, C, D, and E using properties defined by all particles within the dashed boundaries. The key point in all of these examples, is that the baseline halo catalog is used to define the sample of halos whose clustering I study.

Figure 20 shows the results of this investigation. As a reference for the reader, the dashed lines in Fig. 20 reproduce the same MCFs depicted in Fig. 15. The solid lines in Fig. 20 show MCFs using the samples matched to the baseline halo catalogs. To be explicit, in the left panel of Fig. 20 I show the MCFs of all halos that are host halos in the  $\Delta = 20\text{m}$  baseline halo catalog for which the mass and concentration (the mark in this case) of each halo has been computed using the value of  $\Delta$  that corresponds to the color of the curve. The analogous statement is true for the middle and right panels. For both the solid and dashed dark blue lines, corresponding to  $\Delta = 625\text{m}$ , the halos are assigned masses and concentrations using a  $\Delta = 625\text{m}$ . In the case of the solid line, only halos defined as hosts in the  $\Delta = 20\text{m}$  baseline catalog are included in the computation of the MCF. For each halo sample, the baseline values of  $\Delta$  are chosen effectively to remove assembly bias at large scales based on the results in Fig. 15 and Fig. 16. These baseline overdensities are  $\Delta = 20\text{m}$  for the L0125 sample,  $\Delta = 40\text{m}$  for the L0250 sample, and  $\Delta = 250\text{m}$  for the L0500 sample.

Compare pairs of dashed and solid lines at the same  $\Delta$  threshold (same color) in Fig. 20. The difference between a pair of solid and dashed lines at fixed  $\Delta$  is caused entirely by the exclusion of some halos from the lower  $\Delta$  catalog due to a change in halo definition. For the low-mass sample (L0125) and mid-mass sample (L0250), the solid lines exhibit greatly reduced concentration-dependent clustering. I conclude that for relatively low halo masses

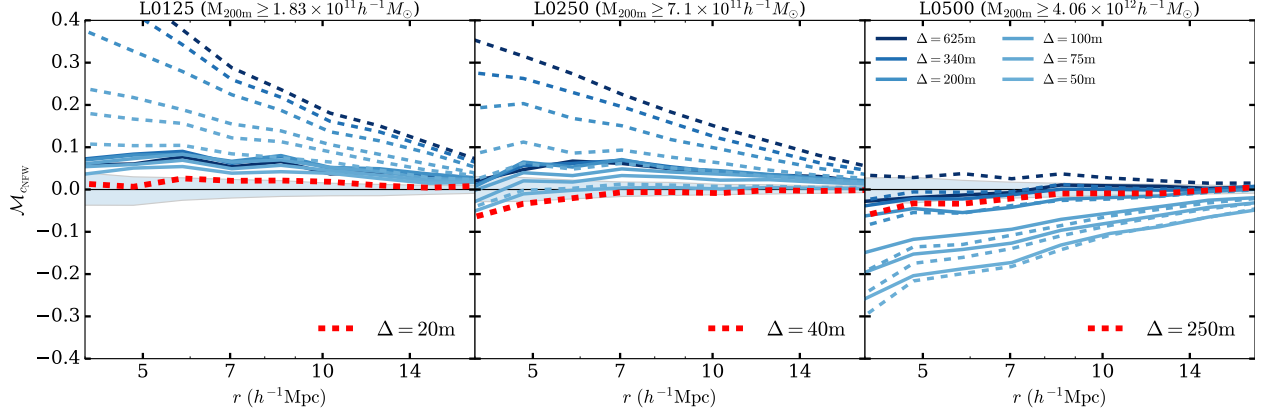


Figure 20 The marked correlation function for the NFW-defined halo concentration parameter matched to baseline halo catalogs. The dashed lines reproduce the MCFs of Fig. 15 for reference. The solid lines plot the marked correlation function using NFW-defined halo concentration as the mark, for a catalog of host halos matched to the host halos in the baseline catalog corresponding to the  $\Delta$  that best mitigates assembly bias (shown as a red dashed line in each panel). For each value of  $\Delta$  only host halos that also appear in the baseline  $\Delta$  halo catalog are used in constructing the MCF shown by the solid lines. For each value of  $\Delta$ , halo properties are determined using the indicated value of  $\Delta$  and *not* the baseline value. The baseline  $\Delta$  only determines which halos are included in the sample. See the main text for further details on sample selection. All lines correspond to different values of  $\Delta$ , with dark blue (light blue) corresponding to  $\Delta = 625\text{m}$  ( $\Delta = 50\text{m}$ ). The left (middle/right) panel shows the results for the L0125 (L0250/L0500) data set utilizing the low mass (mid mass/high mass) cutoffs. The shaded bands represent 2-sigma confidence regions generated by randomization of the marks. This figure demonstrates that most of the assembly bias is removed by the classification of host and satellite halos, rather than by altering the measurement of halo properties such as concentration.

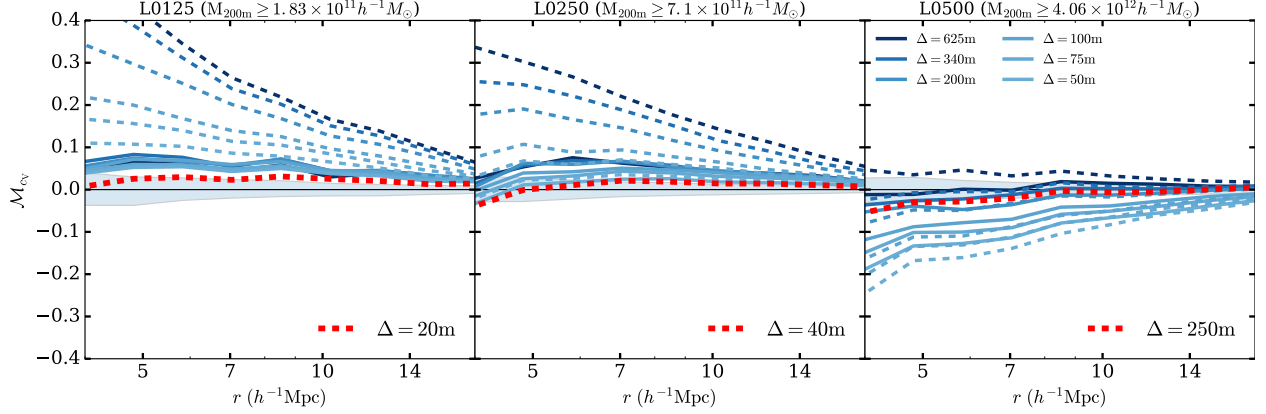


Figure 21 As Fig. 20, with the mark defined as velocity ratio defined concentration.

( $M_{200m} \lesssim 4 \times 10^{12} h^{-1} M_{\odot}$ ), selecting particular halos as hosts eliminates the majority of concentration-dependent clustering. In other words, the majority of the assembly bias signal in this mass range is caused by halos that are nearby neighbors of other, larger halos. At high mass (the L0500 sample), the story is somewhat different. The reclassification of halos as hosts or satellites dependent upon  $\Delta$  makes little difference (dashed and solid lines of the same color are similar). Fig. 21 displays analogous results for the velocity-defined concentration parameter  $c_v$ , suggesting that these results are not driven by any subtle effect of fitting to an NFW density profile.

These results suggest that much of concentration-dependent clustering at relatively low mass ( $M_{200m} \lesssim 4 \times 10^{12} h^{-1} M_{\odot}$ , masses corresponding to other values of  $\Delta$  can be approximated from Fig. 5) is driven by the interactions of nearby halos. They also suggest that subsuming larger regions into halo definitions to accommodate these interactions may be well motivated and practically useful in the context of halo occupation models. It may even be possible to optimize halo definitions for specific applications. Likewise, these results suggest that at high mass ( $M_{200m} \gtrsim 4 \times 10^{12} h^{-1} M_{\odot}$ ), interactions among neighboring halos does not appear to be the predominant cause of concentration-dependent clustering. This is consistent with previous literature suggesting that concentration-dependent clustering may be understood as a property of the initial conditions on large-scales according to the excursion set approach (Zentner, 2007; Dalal et al., 2008) for high-mass halos, but is caused by nonlinear interactions at the low-mass end of the halo mass spectrum (e.g., Wang et al., 2008; Warnick

et al., 2008; Dalal et al., 2008; Hahn et al., 2009; Ludlow et al., 2009; Lacerna & Padilla, 2011; Borzyszkowski et al., 2017).

Lastly, notice that in the left and middle panels of Fig. 20, there is some modest residual assembly bias displayed by the solid lines; the solid lines are not all consistent with zero. This suggests that a small part of the reduction in concentration-dependent clustering is caused by the change in the concentration mark values when halo definition changes. Some of the change may come from the introduction of noise in the concentration measurement that is more weakly correlated with large-scale environment upon redefining halos with larger radii (smaller  $\Delta$ ). Changes in halo masses would also result in changes in the mark values. Nonetheless, for low-mass halos, this residual assembly bias is generally quite modest compared with the concentration-dependent clustering of halos defined with more traditional values of the overdensity threshold.

### 5.3.3 How Does Halo Redefinition “Fail?”

The reader may have noticed that Fig. 18 demonstrates a situation where our change of halo redefinition to smaller values of  $\Delta$  *increases* the measured halo assembly bias. This flies in the face of one of the key goals of this project: finding a halo definition which minimizes halo assembly bias across all halo properties, such that halo modeling can continue unabated with a new definition. The fact that this has the opposite trend as the other halo properties ultimately renders this method impossible across all halo properties; any definition that reduces halo assembly bias with respect to spin will increase halo assembly bias with respect to shape, as both of these marks have a fiducial halo definition with positive spin assembly bias. As a result, the naïve conclusion may be to state that halo redefinition has failed to address the problem at hand.

Rather, I would note that this actually gives us information that we did not previously have before on the problem. As shown in Fig. 5, the change of halo definition roughly corresponds to a change in halo size. There is some noise upon this, as halos in overdense environments need to extend their halo size more significantly to reduce their average density compared to halos in underdense environments which are typically not adding additional

material. The general trend of this behavior allows us to utilize the choice of halo definition as a way to determine *how* halo assembly bias arise; in this way, I move from a primary goal of attempting to mitigate assembly bias to understanding the underlying nature and physics that might lead toward assembly bias.

With regards to this, there are three key trends to interpret, in the context of changing the size of the halo with definition.

1. As the average radius of a halo increases, halos of high concentration become less strongly clustered.
2. As the average radius of a halo increases, halos of high shape become less strongly clustered.
3. As the average radius of a halo increases, halos of high spin become more strongly clustered.

In the case of halo concentration, I have mentioned above there is a distinct mass dependence which is suggestive of concentration dependent clustering being driven by two separate physical processes. However, as seen in Fig. 19, the change in halo definition has a similar impact across a broad range of halo masses. This suggests that some level of halo assembly bias across all masses is driven by halos that are within a few multiples of the  $\Delta = 200$  halo radius. I interpret the result as being driven by the encompassing of neighbor halos as substructure. Since the most clustered halos have higher concentrations *and* halos in overdense regions tend to increase their halo size more, the method of halo redefinition will preferentially change populations in a way that reduces the clustering of high concentration halos. This will leave halos in underdense regions to have their clustering relatively unchanged and (as these are preferentially low concentration) lead to the reversal of the sign of halo assembly bias.

In the case of halo shape, the mass dependence is fairly weak (as seen in Fig. 19 and halo assembly bias exists for all definitions of halo size that I attempt to utilize. As Lee et al. (2017) has shown that halos of low shape parameter preferentially live in underdense environments (where the halo is expected not to have considerable changes in halo size comparatively), I draw attention to the overdense environments that are anticipated to have



higher values of the shape parameter initially. As the size of the halo increases, the inclusion of new halo substructure may lead to the removal of neighboring high shape parameter host halos while simultaneously decreasing the value of the shape parameter (becoming more prolate). The net result is to drive the clustering signal of high shape objects down for objects in large filamentary structures while having a minimal change on halos existing in underdense filaments.

Finally, in the case of halo spin, I again see a weak mass dependence in Fig. 19, but the trend with halo definition is reversed. As [Lee et al. \(2017\)](#) demonstrates, the spin parameter tends to be smaller for halos in the most underdense and the most overdense environments. Yet as I increase the halo size (again, with a preferential size increase in overdense environments), I find that the clustering as a result of the high spin parameter halos increases. I suggest that this result is driven by neighboring halos being converted into subhalos and serving to significantly increase the measured value of the halo spin by having both large halocentric distances and velocities that are preferentially aligned with the spin in filaments. While I cannot confirm this theory from the present data, I intend to look into this further in the future using a version of **ROCKSTAR** tuned to remove substructure from halo property calculations.

## 6.0 HALO SPLASHBACK RADIUS AND ASSEMBLY BIAS

The reader may note that this study carries many similarities in procedure to recent work on the literature focusing on the “splashback radius” (More et al., 2015; Mansfield et al., 2017; Diemer, 2017; Diemer et al., 2017). Indeed, the possibility of a more physically inspired halo definition is a strong motivator behind this work. The splashback radius of a halo is the halo-centric distance where infalling material reaches the apocenter of its first orbit. By analogy with the spherical collapse model (e.g., Fillmore & Goldreich, 1984), this radius separates the halo region, within which material is orbiting the halo and we expect interactions to be important, from material on first infall, for which we expect interactions to be less important. The average overdensity enclosed by the splashback radii of halos is not constant, but varies with halo mass, accretion rate, and redshift (Mansfield et al., 2017; Diemer, 2017; Diemer et al., 2017). I note that, a priori, there is no reason to expect that using splashback radii and masses would mitigate assembly bias. However, investigating the relation between splashback radii and halo assembly bias gains us a better understanding of what may be the underlying cause of the previously observed mitigation.

## 6.1 RESULTS

### 6.1.1 Statistical Comparisons

The work of Diemer et al. (2017) allows for a broad connection between the results of Chapter 5 and the splashback radii, presented as Fig. 22. The blue points show the radii (in units of  $r_{200m}$ ) of halos selected at each of the  $\Delta$  thresholds that minimize concentration-

dependent halo clustering as a function of halo mass,  $M_{200\text{m}}$ . The previous result suggests that halo definition must be a strong function of halo mass in order to mitigate assembly bias. The red line shows the median splashback radius of halos as a function of  $M_{200\text{m}}$  (Diemer et al., 2017) which exhibits a much weaker dependence on halo mass than needed to mitigate assembly bias. This comparison, however, ignores that the red line is merely an average relation with significant scatter. The overdensities enclosed by the splashback radii of individual halos do span a range similar to the overdensities found in this paper, where the overdensity is a strong function of the mass accretion rate. Since mass accretion rates are expected to be correlated with large-scale environment, the overall effect of using splashback radii is hard to predict.

### 6.1.2 SPARTA Comparisons

To address the matter that the splashback radii of individual halos could have significant differences, I make use of SPARTA catalogs provided to me for the L0500 simulation volume. SPARTA, the Subhalo and PARTicle Trajectory Analysis code (Diemer, 2017), allows for the identification of the splashback radius at the level of individual halos (as opposed to the bulk relation that was utilized in the above analysis). Put simply, the algorithm is fairly straightforward. First, individual particles are identified as members of a halo and traced through time in order to identify the apocenter of their first orbit; effectively this is the splashback radius of an individual particle. From the distribution of splashback radius for a given halo, you can define a halo splashback radius from that distribution. I choose to examine two specific splashback radius:  $r_{\text{sp},50\%}$  and  $r_{\text{sp},87\%}$ , where the number represents which percentile in the splashback radius distribution is being used to represent the overall halo. For each of these halo radius definitions, catalogs have been provided to me with host halo identification carried out for that specific definition. As previously seen, changing the distribution of host halos compared to subhalos can have a considerable impact on the resulting halo assembly bias. From this data, I divide the halo distribution into three samples:

1. Host halos as defined by the  $\Delta = 200\text{m}$  halo definition.
2. Host halos as defined by the  $r_{\text{sp},50\%}$  halo definition.

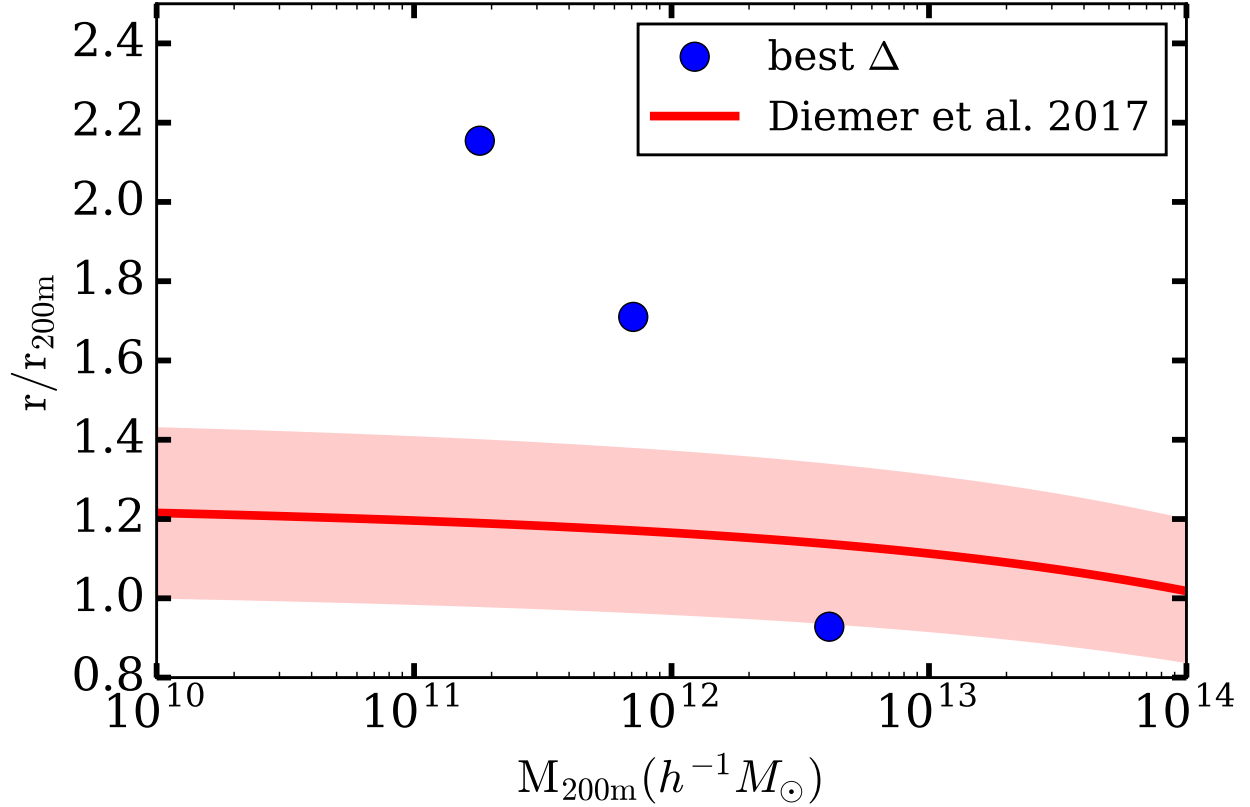


Figure 22 A comparison of the average ratio between  $r_{200m}$  and the splashback radius as determined by [Diemer et al. \(2017\)](#) fitting function (red line) to the average ratio between  $r_{200m}$  and the halo radius determined as our best choice of  $\Delta$  for removal of assembly bias as discussed above (blue circles). The red shaded region represents the 0.07 dex scatter in the Diemer et al. relation. There is some dispersion in  $r_{\Delta}$  but it is quite small (see Fig. 5) so I do not show it on the blue points. Note that the halo mass chosen for the blue points is determined by the mass cutoff in the simulation analysis, as the smallest (and most numerous) halos dominate the calculation.

3. Host halos as defined by the  $r_{\text{sp},87\%}$  halo definition.

Note that the first set of halos only includes those host halos which *also* have identified halo splashback radius; this pushes the sample to the high mass end. In particular, only halos that exceed  $M_{200\text{m}} > 10^{13} h^{-1} M_{\odot}$  are utilized in this analysis. Beyond this, halo properties are calculated using a  $\Delta = 200\text{m}$  spherical aperture and normalized identically to the MCF results above. Any change in the MCF is thus driven by a change in the population of host halos.

I present the results for halo concentration as Fig. 23. The solid lines plot the MCF of halos using the NFW concentration, with each line containing a different halo sample. I emphasize that the only change between these lines is the host halo population; e.g. a splashback radius definition may have less host halos due to the change in size causing a number of nearby host halos to be identified as subhalos. At the given mass range, the fitting relation from Diemer et al. (2017) suggests that the average halo will have a radius roughly 20% larger than  $r_{200\text{m}}$ . Note that even when dealing with each splashback radius being defined for individual halos, the net result is a mild *increase* in the strength of halo assembly bias with respect to halo concentration. This matches the naive anticipation derived from the functional form above.

I also provide the results in Fig. 24 as an example of this relation matching the behavior expected from the mean relation. Note that a large positive halo assembly bias signal is seen for all three halo definitions chosen. Further, the behavior of the splashback radius to reducing halo assembly bias is consistent with the results of Fig. 17. This serves as an additional verification that the halo splashback radius does not appear to mitigate halo assembly bias for either halo concentration or halo shape.

Consider, however, Fig. 25. A first inspection will note that there remains a significant halo assembly bias signal, as to be anticipated for the given mass range given previous results. However, note the minimal impact that a change of halo splashback radius has upon the signal of halo spin driven assembly bias. This contrasts with Fig. 18, where a change in halo definition can lead to a considerable change in the assembly bias signal. This result, while puzzling, may provide some insight into the underlying nature of spin driven assembly bias. Note that halos with large accretion rates are known to have smaller  $r_{\text{sp},50\%}/r_{200\text{m}}$  on

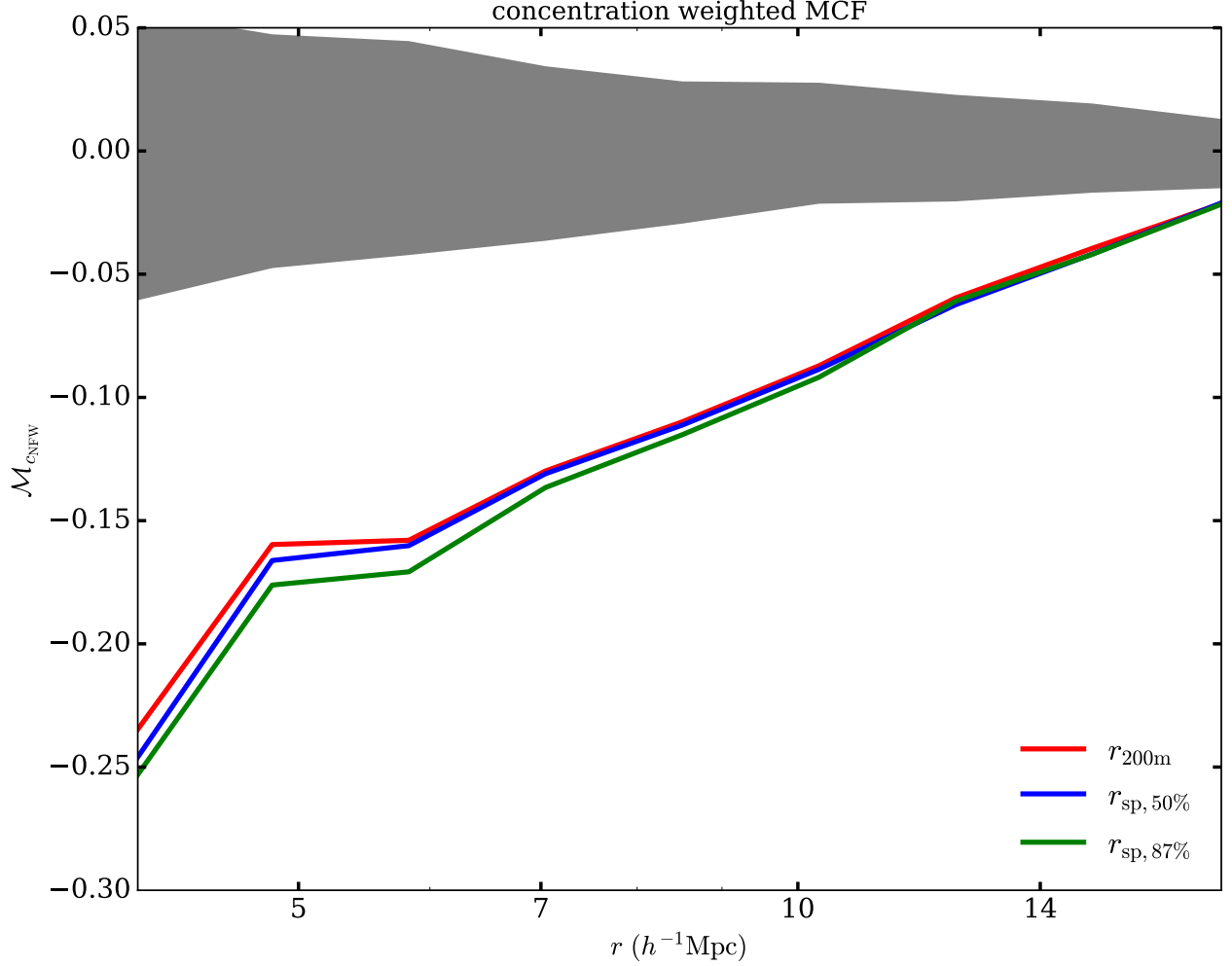


Figure 23 NFW concentration marked correlation functions. The solid lines plot the marked correlation function of halos weighted by NFW concentration for a fiducial  $\Delta = 200\text{m}$  halo definition. The lines correspond to different halo samples, with red corresponding to the fiducial  $\Delta = 200\text{m}$  definition and blue (green) for those halos defined in SPARTA using the 50<sup>th</sup> (87<sup>th</sup>) percentile determined splashback radius. The shaded band contain 98% of 200 MCFs computed from shuffling uniform random marks among the L0500 halo sample.

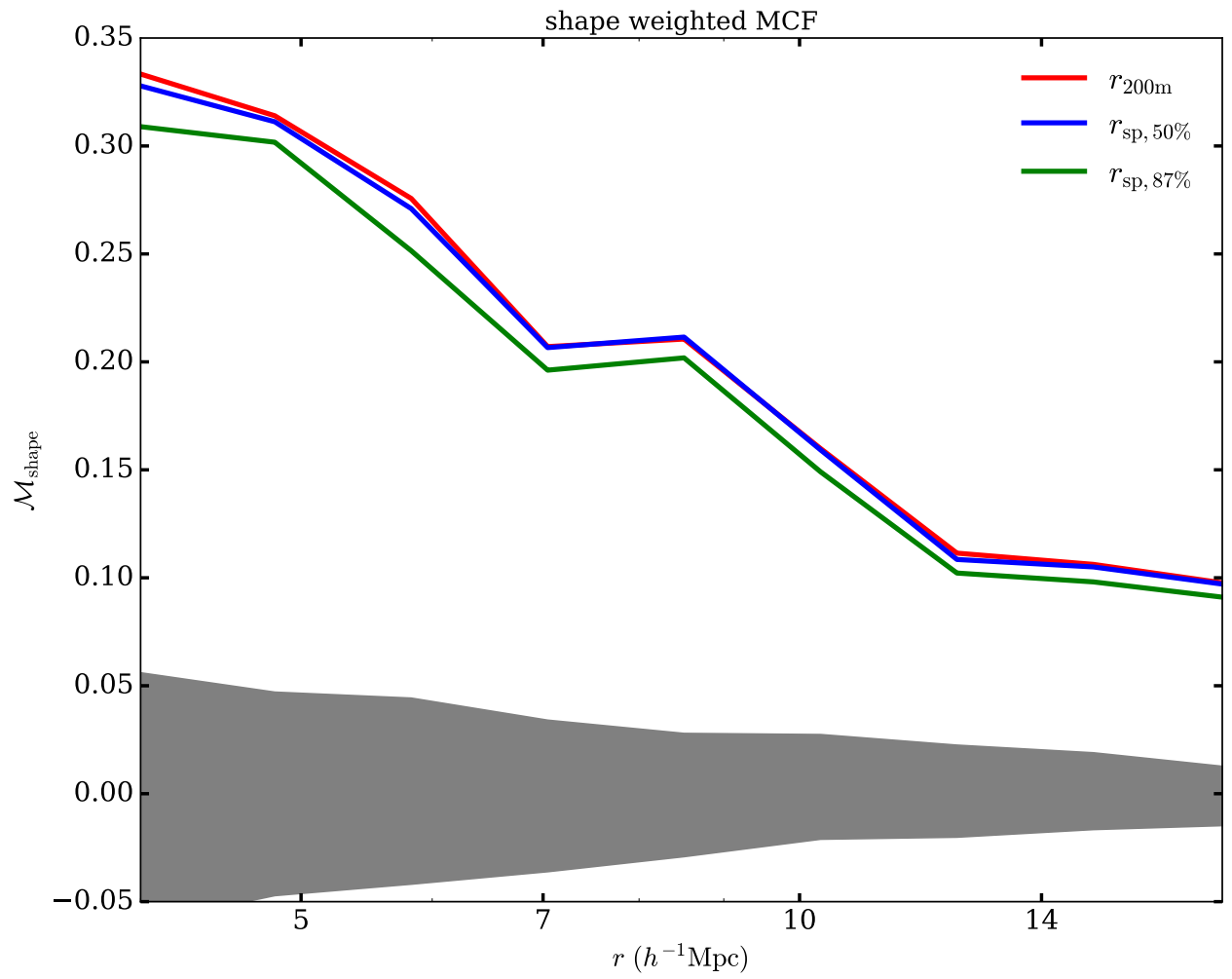


Figure 24 As Fig. 23, using the shape parameter as the property mark.

average. Perhaps those halos with the most halo assembly bias signals do not increase their halo size as aggressively under the splashback radius definition as opposed to the direct halo redefinition carried out in Chapter 5. Examining the cause of this effect is a subject for future work and beyond the scope of this current analysis, as many potential caveats exist to prevent drawing firm conclusions from this particular analysis.



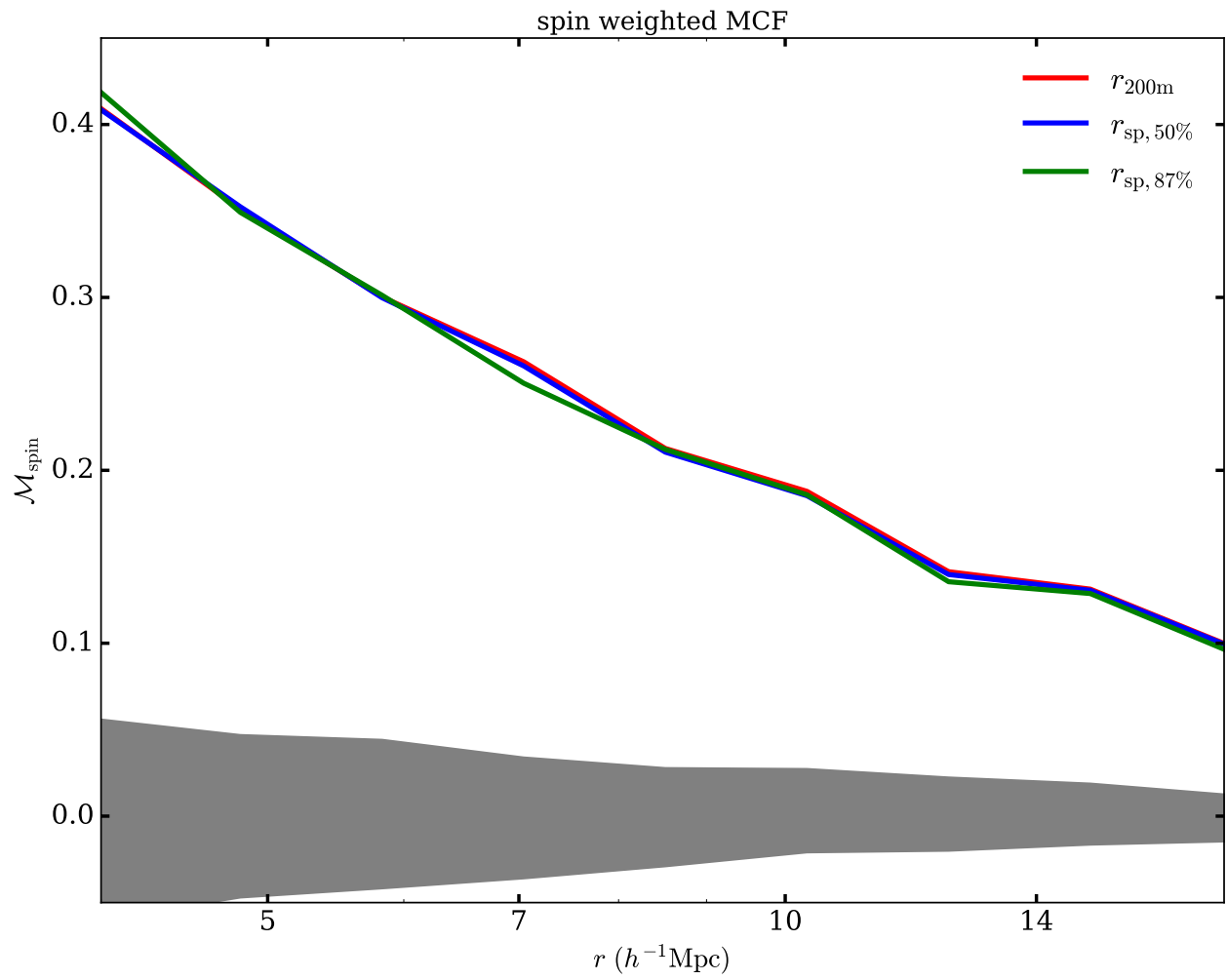


Figure 25 As Fig. 23, using the spin parameter as the property mark.

## 7.0 TECHNICAL DISCUSSION

I have examined the dependence of halo assembly bias upon halo definition, parametrized, for simplicity, by spherical overdensity threshold  $\Delta$ . Fig. 6 presents a pictorial representation of our procedure. This work was motivated as an effort to determine whether or not the dependence of halo clustering strength on halo properties other than mass could be mitigated by judicious choice of halo definition. I further presented results for comparison using the SPARTA splashback radius algorithm, exploring another more physically motivated definition and drew comparisons against my  $\Delta = 200m$  assembly bias measures. I now summarize the results of this analysis below, with a focus on the details for the expert in the field. A more broad view is taken in Chapter 8 below.

- I confirm the results of the literature using the (Diemer & Kravtsov, 2015) simulations regarding halo assembly bias as a function of halo concentration, halo spin, and halo shape (see Fig. 10, Fig. 12, and Fig. 13). I look at the broad mass dependence at a redshift of  $z = 0$  and note the particularly strong mass dependence of halo concentration dependent assembly bias, even when the underlying halo mass-halo concentration relation has been removed.
- I show that the degree to which halo clustering depends upon auxiliary halo properties varies considerably with halo definition. Even among commonly used halo definitions, such as  $\Delta = 625m$  ( $\Delta = 200c$ ),  $\Delta = 340m$ , and  $\Delta = 200m$ , the strength of assembly bias varies considerably with  $\Delta$ . This is particularly true of halo concentration (see Fig. 19).
- A judicious definition of a halo can greatly mitigate concentration-dependent halo clustering (see Fig. 15 and Fig. 16). However, this requires a halo definition that has a

strong mass dependence; a threshold of  $\Delta = 20$  mitigates assembly bias at low masses ( $M_{200\text{m}} \sim 1.8 \times 10^{11} h^{-1} \text{M}_{\odot}$ ) but induces assembly bias of opposite sense at high masses ( $M_{200\text{m}} \gtrsim 4 \times 10^{12} h^{-1} \text{M}_{\odot}$ ). At high masses, an overdensity of  $\Delta \approx 340\text{m}$  results in very weak concentration-dependent clustering.

- At low mass ( $M_{200\text{m}} \lesssim 4 \times 10^{12} h^{-1} \text{M}_{\odot}$  samples from the L0125 and L0250 simulations), concentration dependent clustering is mostly driven by low-mass halos in the immediate neighborhood of larger halos that have had their properties altered through interactions with the larger neighbors. This is not the case for more massive halos (Fig. 20 and Fig. 21). Assembly bias at the high-mass end has a distinct origin.
- Halo shape-dependent clustering is significant over a wide range of halo definitions, but does not exhibit a strong mass dependence for any particular halo definition (Fig. 17 and Fig. 19). While the trend is for shape dependent halo clustering to be weakened as  $\Delta$  is reduced, it cannot be mitigated with any of the halo definitions examined between  $20\text{m} \leq \Delta \leq 625\text{m}$ .
- Halo spin-dependent clustering demonstrates assembly bias that increases weakly with halo mass. Spin-dependent assembly bias can be mitigated with a threshold of  $\Delta \sim 340 - 625\text{m}$  for the lowest masses ( $M_{200\text{m}} \sim 1.8 \times 10^{11} h^{-1} \text{M}_{\odot}$ ), while considerably larger values of  $\Delta$  must be used at higher masses ( $M_{200\text{m}} \gtrsim 4 \times 10^{12} h^{-1} \text{M}_{\odot}$ , see Fig. 18 and Fig. 19).
- Although our study was partly motivated by recent studies of the “splashback radius,” there is no clear connection between our preferred halo definitions and the average splashback radii of halos as a function of halo mass (Fig. 22).
- Similarly, I see no significant mitigation of assembly bias when utilizing individual splashback radii generated through the SPARTA algorithm (See Fig. 23, Fig. 24, and Fig. 25). The changes in the results for halo concentration and halo shape are consistent with a picture in which the splashback radius is larger on average than the  $\Delta = 200\text{m}$  halo definition, which maps well onto our previous results.

- However, I note that those halos with the largest accretion rates (which may be expected to have the most irregular shapes) are those whose halos do not have significantly increased splashback radius. This may be the source for the inconsistency with Fig. 25 with respect to the results of Fig. 18, in which increasing halo size typically increases the halo assembly bias. This suggests a connection between halo accretion rate and halo spin assembly bias.
- I compile many of our the results in Fig. 19, which gives an example of the mass and halo definition ( $\Delta$ ) dependence of the strength of halo assembly bias. This is the first detailed exploration of halo definition impacting assembly bias in the literature.

I conclude that a single, simple, definition of halo size based on overdensity or a similar criterion cannot, by itself, be exploited as a method to mitigate halo assembly bias. At the very least, mitigation requires any such halo definition to be mass dependent, and likely requires halo definitions that are considerably more complicated than what I have explored here. It is possible that these strategies may be more fruitful when applied to halo properties that have not studied in the present work. A prominent example of such a property would be a measure of halo formation history, though this potentially raises the computational cost due to needing to trace through the halo merger history in some form. A thorough examination of other potential properties may still merit further analysis. Nonetheless, as more and more precise galaxy clustering data become available, we must continue to seek tools that may be used to interpret such high-quality data. Revisiting and reconsidering the concept of a dark matter halo may continue to be one aspect of this search.

The results of this study offer several promising extensions in the near immediate future, to be resolved in future publications. The most straightforward analysis I intend to explore is to continue utilizing the **ROCKSTAR** halo finder in order to confirm the hypothetical models I have made to explain the current results. For example, the previous analysis has proposed that halo properties can become modified by the inclusion of new substructure inside of the expanding host halo. I currently have access to a modified version of **ROCKSTAR** which explicitly excludes subhalos from the property analysis; it is possible that if the inclusion of subhalos is the fundamental driver of halo assembly bias, halo properties calculated

in the absence of subhalo contamination should be more robust. Another possibility in the search for a more robust halo definition is to attempt a halo definition dependent upon the local environment; this could vary from defining halo definitions with respect to some aperture centered upon the density peak rather than the total background of the universe to using this same aperture to choose different halo definitions for different local environments. Testing this would be more computationally intensive due to needing to determine how various apertures for determining the local density and other similar assumptions would impact the result; there may also be nontrivial coding issues regarding the parallelization of the **ROCKSTAR** algorithm. The latter method bears some similarities to the suggestion of a mass based halo definition, in which a halo could be defined  $M_{200m}$ , which is then utilized to choose a more optimal halo definition based on the results in Fig. 19. The ultimate goal would be to attempt to find more self-similar halos, though the different dependences as a function of halo property still suggests that no global solution will be found to fit *all* halo parameters.

I reiterate that one of the major take-aways from this study is the fact that the impact of halo definition on the measured assembly bias is comparable to the impact from the choice of halo mass observed. The results of this study then serve to help quantify this effect for ease of comparison between different studies. This is especially important for those attempts that seek to measure the level of halo assembly bias in observational data; the difference between choice of halo definition, choice of halo mass, and halo properties examined can potentially conspire to lead one to measure zero assembly bias. Extra care ultimately needs to be taken to make a true “apples-to-apples” comparison between results in the literature. Note that I have not explored at this time the potential impact of other halo definition choices *other* than spherical overdensity halos; careful exploration may be necessary of options such as FOF halos as well. One can easily imagine a universe where different choices of the linking length can have an impact on the measured assembly bias, under motivation similar to that of this work for expanding the halo radius.

To summarize the technical discussion, the above analysis finds the following major effects that need to be accounted for in future halo analysis. Halo concentration driven assembly bias is a strong function of both halo mass and halo definition; while best choice

halo definitions to mitigate assembly bias are possible for this parameter, they are not universal across all halo mass possibilities. Halo shape and halo spin have assembly bias that is a function of halo definition, but only a weak function of halo mass. This makes these properties better candidates for halo redefinition, though the impact of changing halo size is different in size for each; increasing halo size tends to increase halo spin assembly bias and decrease halo shape assembly bias. Finally, the splashback radius halo definition does not demonstrate any significant preference over our naïve method of halo redefinition to justify the greatly increased computational load. It does, however, give teasing hints toward the understanding of the origin of halo spin assembly bias, however. Finally, future work can focus on the continued exploration of halo definition choices, the choice to include or exclude substructure in halo modeling, and improving the ability to make proper comparisons within the literature.

The next chapter shall provide more a more broad, all-encompassing view of the potential for this research to impact the field as a whole, without necessarily becoming bogged down in a technical morass. I will present many possibilities based off these technical results which have the potential to become entire dissertations in their own right; while some of these projects may continue with my personal touch, the sheer breadth of what halo redefinition can offer to the field necessitates working on future graduate students to truly explore all the potential.

## 8.0 CONCLUSIONS

The previous chapter of discussion focused primarily on the more technical aspects of our results from the perspective of an expert in the field. However, it should be demonstrated that the results of this study have broad reach across the field of both cosmology and galaxy evolution. In this chapter, I present several key fields that the results of this study will impact, what current experiments or surveys will likely need to account for these methods and results, and what possibilities for future work exist. The first major thing to consider is comparison within the literature, with a particular focus on the attempts to detect halo assembly bias; this study allows an entirely new means by which to ascertain that comparisons are being done in a robust way. There is also the broad sense in which halo assembly bias can serve as a probe on the nature of galaxy formation and evolution. Finally, an important consideration to consider is the application to improving halo model driven calculations, where halo assembly bias can be a potential source of systematic errors. This final case has broad impact to cosmological parameter estimation. I discuss possible applications of these results to each of these major considerations and possible avenues for future study.

Let us first examine the case of comparison within the literature. As has been shown throughout this work (most prominently in Fig. 19), halo definition and halo mass are *both* important parameters to the understanding of halo assembly bias. In fact, after the removal of any underlying halo mass-halo shape or halo mass-halo spin relations, halo definition impacts the measure of halo assembly bias more significantly than halo mass. Consider that the comparison between  $\Delta = 200m$  and  $\Delta = 625m$  is approximately the difference between  $\Delta = 200m$  and  $\Delta = 200c$ ; this is the level of difference that is common even within work on halo assembly bias. The difference can be significantly greater in other fields! In the context of reproducibility, it is clear that neglecting clear definition of the process of halo finding

is a serious problem. Clarity is necessary for the field to be able to work on this problem adequately.

For the future of this question though, further work needs to be carried out. An observational detection of halo assembly bias necessitates two major changes in the approach laid out here utilizing only simulations. The first is being able to connect halo properties to observables in the universe; mapping out halo concentrations in the observable universe is too difficult for the raw statistical power that is necessary for being able to pull out halo assembly bias. As such, proxies for halo concentration (or halo formation time) are utilized in typical analysis. A second change of approach is the matter of satellite contamination. In this work, it is a trivial step to separate host halos from subhalos and substructure. In observation, this can be a potentially dominant source of error. One future extension to our work will be to apply these same physical procedures to catalogs created with known halo assembly bias and test how the ability to parameterize this assembly bias changes as a function of halo definition. This may determine that some potential halo definitions are less sensitive to effects such as satellite contamination.

I opened this study with a key question: “can a halo definition be chosen such that there is no halo assembly bias across multiple halo properties.” Our answer is there exists no clear solution to the problem of halo assembly bias. In a way, this work serves to recontextualize the problem of halo assembly bias. The fact that halo assembly bias is likely unable to be removed through halo redefinition as a result of the complex and differing halo mass and halo definition dependences necessitates a change of approach. Rather than a sharp focus on the removal of halo assembly bias, we now shift to a desire to understand halo assembly bias in the greater context of galaxy formation and galaxy clustering. Instead of considering it a potential nuisance, it is likely time to shift to models which explore halo assembly bias as a potential marker to teach us about how the halo-galaxy connection depends upon other parameters. For example, the signal of halo assembly bias with respect to halo spin seen in Fig. 18 implies that halo spin is intrinsically tied to large halocentric distances. As halo spin is known to be connected to galaxy spin (Jiang et al., 2018), this may imply that the source of galaxy spin should also be motivated by physics on large halocentric distances. Thus, halo assembly bias can now motivate our study of galaxy formation and



evolution.

A natural extension then is to consider how we can leverage this new connection. The first possibility is to consider the direction of moving from halo catalog to a mock galaxy catalog, utilizing different methods of halo redefinition. As host halos and substructure are frequently treated differently in these halo models, the resulting mock galaxy catalog should simultaneously change as a function of halo redefinition. Consider tracking a statistic to compare the mock galaxy catalog to actual observations at realizing the same populations, such as the expected clustering of galaxies. By seeing how this changes as a function of halo redefinition, one can attempt to probe where specific parts of the model are breaking down. For example, if halo definitions are chosen where galaxy spin assembly bias should be minimal, but the observation still demonstrates an excess clustering, it suggests that galaxy spin assembly bias is driven through what is now host and subhalo interactions. Our models can then be iterated upon in order to determine what scales matter for which parameters and which sections are more important to subhalos versus hosts. While this work may be of considerable technical detail, it provides an interesting new outlook on the field.

Of course, the reverse situation is also possible. Rather than considering how we can move from halos to galaxies, we can take the opposite approach. Working with smaller halo definitions than normal may allow us to avoid some of the possible fear of satellite contamination. A confined halo definition can be used in order to relate galaxy clustering properties back to their underlying halos with less risk, due to the relatively small halos necessitating that other galaxies be within their own host halos. In this case, our work paves the way for the necessary corrections that allow us to relate back to definitions that we are more familiar with. We can assign halos of one definition and then imagine a situation in which one uses relations such as the relative assembly bias and relative masses to transition from a collection of very small halos to a collection of larger and interacting halos more in line with our expectations. In this case, using a very tightly confined halo serves as a temporary working definition.

In the immediate future, we will remain motivated by a need to determine accurate cosmological parameters; as discussed in Section 2.2, the literature suggests that not accounting for halo assembly bias correct will lead to a biasing of inferences about the behavior of

dark energy. This work gives a stronger understanding of how halo definition changes the measured assembly bias; this will be of critical importance to combining different probes for the evolution of the universe, which can often have drastically different choices in halo definition. Even within a single cosmological probe, there are exciting opportunities for research with halo redefinition. I lay out several possibilities below.

One possible analysis is that of cluster count cosmology. The idea here is straightforward: by counting the number of galaxy clusters of a given total mass, you can connect to expectations from a given cosmology (Campanelli et al., 2012, e.g.). Naturally, as dark matter is the dominant source of mass, this makes the galaxy-halo connection an important point of consideration. Further, multiwavelength cluster cosmology makes use of both the  $\Delta = 200m$  and  $\Delta = 500m$  definitions within their analysis, which I have shown have significantly different response to halo assembly bias. While it has not yet been studied *how much* of an impact that neglecting the change in assembly bias as a function of halo definition may have on such analysis, this work strongly motivates a need to treat such choices less trivially. In a more subtle manner, the fact that a change in halo definition can change host (and subhalo) occupations can have an impact in the semianalytic models that are used to generate galaxy mock catalogs for testing and verification as satellite merger is often an important component in this modeling (Benson, 2012, e.g.).

For another case, consider the recently completed Dark Energy Survey. With a large number of galaxies and a measure of how their shapes are distorted due to the gravitational interactions of intervening matter (referred to as shear in the literature), the collaboration was able to put excellent constraints on the combination of  $\sigma_8$ , the normalization of the power spectrum in the early universe, and  $\Omega_m$ , the energy density of matter in the universe (DES Collaboration et al., 2017). However, it has been shown in the literature that extending these sorts of analysis to smaller scales could yield significantly better constraints. Understanding halo assembly bias is critical to this future analysis. Note that I have demonstrated that significant assembly bias exists in our sample for a separation range ranging between  $5-10h^{-1}$  Mpc; at these separations, the strength of nonlinear effects has led to the breakdown of perturbation theory approaches (Carlson et al., 2009). Understanding assembly bias in simulations is thus a key element to future surveys being able to improve their constraints;

specific work should be done to examine just how much potential bias can be introduced if assembly bias is present in the data, but unaccounted for in the analysis.

One final case for cosmological analysis worth considering is that of intrinsic alignments. When carrying out the measurement of galaxy shapes briefly mentioned above, the simplest analysis is to assume that all galaxy shapes are randomly distributed initially, but intervening mass can create alignments as their shapes are distorted. However, with the hierarchical growth of structure in the universe coupled with features such as the cosmic web there is the strong possibility of an intrinsic alignment between galaxies; that is to say they have some preferential alignment with neighbors rather than being randomly distributed (Blazek et al., 2016). As different galaxy populations are known to have different intrinsic alignment signals (Blazek et al., 2015), it becomes more important to have a strong sense of the galaxy spin-halo spin connection - and the halo spin driven assembly bias is significant. While it is possible that one may approach this problem with the use of new halo definitions (as suggested in our work), it is also quite likely that simply a better understanding of relevant scales, relation to halo definition and halo size, and accounting for assembly bias may be the best approach.

Ultimately, it becomes clear very quickly that halo redefinition can give us exciting new leverage over many interesting and distinct fields today. If somebody needs to make comparisons within the literature, an understanding of halo definition's impact on halo assembly bias is a necessity to making fair and accurate meta-analysis. As a tool for understanding galaxy formation and evolution, halo redefinition gives us a new way to explore the relevant scales of interest and test how well our models connect to the observable universe. Finally, as new surveys begin to focus their analysis in an effort to get a better understanding of the universe around us, halo redefinition helps to give both an understanding of the halo assembly bias that must be accounted for when extending to smaller scales *AND* potentially serves as a mitigation technique for subsets of the analysis. The work that remains is vast and will yield exciting new insights into the world around us.

## BIBLIOGRAPHY

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *Physical Review Letters*, 119, 161101
- Allgood, B., Flores, R. A., Primack, J. R., et al. 2006, *MNRAS*, 367, 1781
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, *ApJ*, 770, 57
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013b, *ApJ*, 762, 109
- Beisbart, C., & Kerscher, M. 2000, *ApJ*, 545, 6
- Benson, A. J. 2012, *New A*, 17, 175
- Berlind, A. A., & Weinberg, D. H. 2002, *ApJ*, 575, 587
- Berlind, A. A., Weinberg, D. H., Benson, A. J., et al. 2003, *ApJ*, 593, 1
- Bertone, G., Hooper, D., & Silk, J. 2005, *Phys. Rep.*, 405, 279
- Bett, P., Eke, V., Frenk, C. S., et al. 2007, *MNRAS*, 376, 215
- Blazek, J., Seljak, U., & Mandelbaum, R. 2016, in *IAU Symposium*, Vol. 308, *The Zeldovich Universe: Genesis and Growth of the Cosmic Web*, ed. R. van de Weygaert, S. Shandarin, E. Saar, & J. Einasto, 452–455
- Blazek, J., Vlah, Z., & Seljak, U. 2015, *J. Cosmology Astropart. Phys.*, 8, 015
- Bode, P., Ostriker, J. P., & Turok, N. 2001, *ApJ*, 556, 93
- Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, *ApJ*, 379, 440
- Borzyszkowski, M., Porciani, C., Romano-Díaz, E., & Garaldi, E. 2017, *MNRAS*, 469, 594
- Bosma, A. 1978, PhD thesis, PhD Thesis, Groningen Univ., (1978)
- Bower, R. G. 1991, *MNRAS*, 248, 332
- Bryan, G. L., & Norman, M. L. 1998, *ApJ*, 495, 80
- Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, *MNRAS*, 321, 559

- Bullock, J. S., Wechsler, R. H., & Somerville, R. S. 2002, MNRAS, 329, 246
- Campanelli, L., Fogli, G. L., Kahniashvili, T., Marrone, A., & Ratra, B. 2012, European Physical Journal C, 72, 2218
- Carlson, J., White, M., & Padmanabhan, N. 2009, Phys. Rev. D, 80, 043531
- Chaves-Montero, J., Angulo, R. E., Schaye, J., et al. 2016, MNRAS, 460, 3100
- Clifton, T., Ferreira, P. G., Padilla, A., & Skordis, C. 2012, Phys. Rep., 513, 1
- Cole, S., & Lacey, C. 1996, MNRAS, 281, 716
- Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
- Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, MNRAS, 373, 369
- Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303
- Croton, D. J., Stevens, A. R. H., Tonini, C., et al. 2016, ApJS, 222, 22
- Dalal, N., White, M., Bond, J. R., & Shirokov, A. 2008, ApJ, 687, 12
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
- de Blok, W. J. G. 2010, Advances in Astronomy, 2010, 789293
- DES Collaboration, Abbott, T. M. C., Abdalla, F. B., et al. 2017, ArXiv e-prints, arXiv:1708.01530
- Desmond, H., & Wechsler, R. H. 2017, MNRAS, 465, 820
- Diemer, B. 2017, ApJS, 231, 5
- Diemer, B., & Kravtsov, A. V. 2014, ApJ, 789, 1
- . 2015, ApJ, 799, 108
- Diemer, B., Mansfield, P., Kravtsov, A. V., & More, S. 2017, ApJ, 843, 140
- Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64
- Efstathiou, G., Kaiser, N., Saunders, W., et al. 1990, MNRAS, 247, 10P
- Eifler, T., Krause, E., Dodelson, S., et al. 2015, MNRAS, 454, 2451
- Faltenbacher, A., Gottlöber, S., Kerscher, M., & Müller, V. 2002, A&A, 395, 1
- Faltenbacher, A., & White, S. D. M. 2010, ApJ, 708, 469
- Feng, Y., Di-Matteo, T., Croft, R. A., et al. 2016, MNRAS, 455, 2778

- Fillmore, J. A., & Goldreich, P. 1984, *ApJ*, 281, 1
- Fisher, J. D., & Faltenbacher, A. 2016, ArXiv e-prints, arXiv:1603.06955
- Freedman, W. L. 2017, *Nature Astronomy*, 1, 0169
- Freeman, K. C. 1970, *ApJ*, 160, 811
- Gallart, C., Monelli, M., Mayer, L., et al. 2015, *ApJ*, 811, L18
- Gao, L., & White, S. D. M. 2007, *MNRAS*, 377, L5
- Gao, L., White, S. D. M., Jenkins, A., Frenk, C. S., & Springel, V. 2005, *MNRAS*, 363, 379
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *Space Sci. Rev.*, 123, 485
- Garrett, K., & Dūda, G. 2011, *Advances in Astronomy*, 2011, 968283
- Gott, III, J. R., Jurić, M., Schlegel, D., et al. 2005, *ApJ*, 624, 463
- Götz, M., & Sommer-Larsen, J. 2003, *Ap&SS*, 284, 341
- Gunn, J. E., & Gott, III, J. R. 1972, *ApJ*, 176, 1
- Hahn, O., Carollo, C. M., Porciani, C., & Dekel, A. 2007a, *MNRAS*, 381, 41
- Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007b, *MNRAS*, 375, 489
- Hahn, O., Porciani, C., Dekel, A., & Carollo, C. M. 2009, *MNRAS*, 398, 1742
- Han, J., Li, Y., Jing, Y., et al. 2018, ArXiv e-prints, arXiv:1802.09177
- Harker, G., Cole, S., Helly, J., Frenk, C., & Jenkins, A. 2006, *MNRAS*, 367, 1039
- Hearin, A. P., Watson, D. F., Becker, M. R., et al. 2014, *MNRAS*, 444, 729
- Hearin, A. P., Zentner, A. R., Berlind, A. A., & Newman, J. A. 2013, *MNRAS*, 433, 659
- Hester, J. A., & Tasitsiomi, A. 2010, *ApJ*, 715, 342
- Holmberg, E. 1941, *ApJ*, 94, 385
- Hubble, E. P. 1929, *ApJ*, 69, doi:10.1086/143167
- Inoue, K. T., Takahashi, R., Takahashi, T., & Ishiyama, T. 2015, *MNRAS*, 448, 2704
- Jiang, F., Dekel, A., Kneller, O., et al. 2018, ArXiv e-prints, arXiv:1804.07306
- Jing, Y. P., & Suto, Y. 2002, *ApJ*, 574, 538
- Katz, N. 1992, *ApJ*, 391, 502

- Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201
- Klypin, A., & Holtzman, J. 1997, ArXiv Astrophysics e-prints, astro-ph/9712217
- Klypin, A., Prada, F., Yepes, G., Hess, S., & Gottlober, S. 2013, ArXiv e-prints, arXiv:1310.3740
- Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
- Knebe, A., Knollmann, S. R., Muldrew, S. I., et al. 2011, MNRAS, 415, 2293
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, ApJ, 609, 35
- Lacerna, I., & Padilla, N. 2011, MNRAS, 412, 1283
- . 2012, MNRAS, 426, L26
- Lacey, C., & Cole, S. 1993, MNRAS, 262, 627
- . 1994, MNRAS, 271, 676
- Lee, C. T., Primack, J. R., Behroozi, P., et al. 2017, MNRAS, 466, 3834
- Lewis, A., & Bridle, S. 2002, Phys. Rev., D66, 103511
- Li, Y., Mo, H. J., & Gao, L. 2008, MNRAS, 389, 1419
- Linke, L., Schwinn, J., & Bartelmann, M. 2017, ArXiv e-prints, arXiv:1712.04461
- LSST Dark Energy Science Collaboration. 2012, ArXiv e-prints, arXiv:1211.0310
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201
- Ludlow, A. D., Navarro, J. F., Springel, V., et al. 2009, ApJ, 692, 931
- Macciò, A. V., Dutton, A. A., van den Bosch, F. C., et al. 2007, MNRAS, 378, 55
- Mansfield, P., Kravtsov, A. V., & Diemer, B. 2017, ApJ, 841, 34
- Mao, Y.-Y., Williamson, M., & Wechsler, R. H. 2015a, ApJ, 810, 21
- . 2015b, ApJ, 810, 21
- McClintock, T., Rozo, E., Becker, M. R., et al. 2018, ArXiv e-prints, arXiv:1804.05866
- Miyatake, H., More, S., Takada, M., et al. 2016, Physical Review Letters, 116, 041301
- Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

- Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
- More, S., Diemer, B., & Kravtsov, A. V. 2015, ApJ, 810, 36
- Mortonson, M. J., Weinberg, D. H., & White, M. 2014, ArXiv e-prints, arXiv:1401.0046
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
- Nelson, D., Pillepich, A., Genel, S., et al. 2015, Astronomy and Computing, 13, 12
- Onions, J., Knebe, A., Pearce, F. R., et al. 2012, MNRAS, 423, 1200
- Paranjape, A., Hahn, O., & Sheth, R. K. 2018, MNRAS, 476, 3631
- Park, C., Choi, Y.-Y., Kim, J., et al. 2012, ApJ, 759, L7
- Peebles, P. J. E. 1969, ApJ, 155, 393
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16
- . 2016, A&A, 594, A13
- Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2012, MNRAS, 423, 3018
- Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
- Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1980, ApJ, 238, 471
- Scannapieco, C., Wadepuhl, M., Parry, O. H., et al. 2012, MNRAS, 423, 1726
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
- Sheth, R. K. 2005, MNRAS, 364, 796
- Sheth, R. K., & Tormen, G. 2004, MNRAS, 350, 1385
- Skibba, R., Sheth, R. K., Connolly, A. J., & Scranton, R. 2006, MNRAS, 369, 68
- Smith, M. C., Sijacki, D., & Shen, S. 2018, MNRAS, 478, 302
- Smith, S. 1936, ApJ, 83, 23
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints, arXiv:1503.03757
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
- Sunayama, T., Hearin, A. P., Padmanabhan, N., & Leauthaud, A. 2016, MNRAS, 458, 1510



- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, 688, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, 724, 878
- Turner, M. S., Steigman, G., & Krauss, L. M. 1984, *Physical Review Letters*, 52, 2090
- van Daalen, M. P., Angulo, R. E., & White, S. D. M. 2012, *MNRAS*, 424, 2954
- van Daalen, M. P., & Schaye, J. 2015, *MNRAS*, 452, 2247
- Vogele, M. S., Hoyle, F., Rojas, R. R., & Goldberg, D. M. 2004, in *IAU Colloq. 195: Outskirts of Galaxy Clusters: Intense Life in the Suburbs*, ed. A. Diaferio, 5–11
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Wang, H., Mo, H. J., Chen, S., et al. 2018, *ApJ*, 852, 31
- Wang, Y., Yang, X., Mo, H. J., et al. 2008, *ApJ*, 687, 919
- Warnick, K., Knebe, A., & Power, C. 2008, *MNRAS*, 385, 1859
- Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, *ApJ*, 568, 52
- Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, *ApJ*, 652, 71
- White, S. D. M., & Rees, M. J. 1978, *MNRAS*, 183, 341
- Wu, H.-Y., Rozo, E., & Wechsler, R. H. 2008, *ApJ*, 688, 729
- Yèche, C., Palanque-Delabrouille, N., Baur, J., & du Mas des Bourboux, H. 2017, *J. Cosmology Astropart. Phys.*, 6, 047
- Zentner, A. R. 2007, *International Journal of Modern Physics D*, 16, 763
- Zhao, D. H., Jing, Y. P., Mo, H. J., & Börner, G. 2009, *ApJ*, 707, 354
- Zhao, D. H., Mo, H. J., Jing, Y. P., & Börner, G. 2003, *MNRAS*, 339, 12
- Zheng, Z., Coil, A. L., & Zehavi, I. 2007, *ApJ*, 667, 760
- Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, *ApJ*, 633, 791
- Zwicky, F. 1933, *Helvetica Physica Acta*, 6, 110